

# Causes and consequences of adult sepsis in Blantyre, Malawi

-

Thesis submitted in accordance with the requirements of the Liverpool School of Tropical Medicine for the degree of Doctor in Philosophy by Joseph Michael Lewis

September 2019



*For Nell,  
Aoife,  
and Una.*





# Collaborators and co-workers

This thesis is the result of my own work. In some cases, work was done in collaboration with others. The roles of collaborators and co-workers in this project are detailed here.

Screening, enrolment and follow up of study participants: Research nurses Lucy Keyala, Tusekile Phiri, Grace Mwaminawa, field workers Witness Mtambo and Gladys Namacha.

Logistic support and sample portering: Monica Matola.

Laboratory processing of stool, rectal swab, serum, urine and whole blood samples, stool culture, urinary LAM testing and DNA extractions: Madalitso Mphasa and Rachel Banda.

Building and maintaining study databases and electronic data capture forms: MLW data department but especially Lumbani Makhaza

Diagnostic bench microbiology of blood and CSF culture: MLW Microbiology Department.

Sputum Xpert testing and mycobacterial blood culture: Malawi College of Medicine TB laboratory.

Library preparation, whole genome sequencing, post sequencing archiving: Wellcome Sanger Institute pipeline and pathogen informatics team.

Bioinformatics training and support with bioinformatic analysis: Mat Beale and Nick Thomson at the Wellcome Sanger Institute

Statistical support: Brian Faragher and Marc Henrion.



# Abstract

Sepsis, defined as a life-threatening organ dysfunction triggered by infection, carries a high mortality. Recent improvements in outcome high-income settings have been driven by prompt antimicrobial therapy and fluid resuscitation but mortality remains disproportionately high in low-resource settings like the nations of sub-Saharan Africa (sSA). Sepsis therapy here often consists of empiric, prolonged courses of broad-spectrum antimicrobials, especially third generation cephalosporins like ceftriaxone, which may be driving the rise of ceftriaxone-resistant extended-spectrum  $\beta$ -lactamase producing Enterobacteriaceae (ESBL-E). However the aetiology of sepsis in sSA is far from clear, and in this thesis I hypothesise that it may be possible to improve outcomes in sepsis whilst reducing selection pressure for ESBL-E, with novel, targeted, antimicrobial strategies tailored to the pathogens that are truly causing sepsis here.

To that end, I present findings from a clinical cohort study of sepsis in Blantyre, Malawi, with two aims: first, a description of the presentation and outcomes of sepsis in Blantyre, with a focus on aetiology and an analysis of the determinants of mortality; and secondly, a description of the gut mucosal carriage of ESBL-E in sepsis survivors (as well as antibiotic unexposed inpatient and community controls) as they pass through the hospital to identify determinants of carriage. An expanded package of diagnostic tests was used to define sepsis aetiology, and serial stool sampling with selective culture for ESBL-E used to define ESBL-E carriage. I use whole-genome sequencing of cultured ESBL *E. coli* to track bacteria and mobile genetic elements within participants over time, and continuous time Markov models to provide insight into the drivers of carriage.

I find that the majority of participants with sepsis are young, and HIV-infected. Disseminated tuberculosis (TB) dominates as a cause of sepsis, and there is an association of receipt of antituberculous chemotherapy with survival that suggests an expanded role for TB therapy in these very unwell patients may be beneficial. Sepsis mortality seems to have improved compared to historic cohorts, but post 28-day mortality in HIV-infected individuals is significant.

At baseline gut mucosal ESBL-E carriage is common, with cultured ESBL-E present in the stool of 49% of participants with sepsis on the day of admission. There is further

rapid increase in colonisation prevalence following admission and antibacterial exposure. Associations of baseline colonisation - household crowding and unprotected water sources - suggest both within-household and environmental routes of transmission are important. Genomic analysis suggest unrestricted mixing of ESBL *E. coli* at multiple spatial levels and rapid turnover within the individual, perhaps suggestive of frequent re-exposure. By using the genetic environment of ESBL genes as a proxy for mobile genetic elements (which are difficult to assemble from short read sequencing) I show that, within individuals, the *E. coli* strain-mobile genetic element combination is conserved over time whereas the strain or mobile genetic element alone is not; this suggests that the unit of transmission of ESBL gene to study participants is the bacterium, rather than mobile genetic element.

Longitudinal modelling provides further insight into ESBL-E carriage dynamics: hospitalisation and antibacterial exposure act synergistically to bring about rapid and prolonged carriage driven, in part, by a significant post-antibiotic effect. This effect means that antibacterials act to prolong carriage long after antibacterial exposure stops. In terms of ESBL-E carriage, short courses of antibacterials have a similar effect to longer courses, such that the data generated in this study do not support my hypothesis and it may not be possible to reduce ESBL-E carriage by truncating courses of ceftriaxone. Nevertheless, the post-antibiotic effect deserves further scrutiny to understand the mechanism and as a potential therapeutic target. In addition, the modelling approach suggests cotrimoxazole preventative therapy (CPT) may be a significant driver of long-term ESBL-E carriage, and I suggest that a more nuanced approach to its deployment may be necessary in an era of increasing Gram-negative resistance.

# Acknowledgements

This work would not have been possible without the assistance of many people, not least the participants and guardians who agreed to take part. These people agreed to participate in research in the context of critical illness, and in what follows it must not be forgotten that many of them did not survive: I thank them all.

I owe thanks to many people at MLW. The study team worked incredibly hard and were a joy to work with: Witness Mtambo, Gladys Namacha, Lucky Keyala, Grace Mwimaniwa, Tusekile Phiri, Madalistso Mphasa, Rachel Banda and Monica Matola. Administering the study would have been impossible without the help of Lumbani Makhaza, Linley Mvula and Clemens Masesa in the MLW data department. I am grateful to the support of the MLW microbiology laboratory and Brigitte Denis in particular.

Numerous colleagues gave time to explain, or discuss (or both) multiple ideas that made their way in to this thesis: Jamie Rylance, David Barr, Rebecca Lester, Adam Roberts, Eva Heinz, Brian Faragher, Marc Henrion and Chris Jewell. Jamie generously shared resources from studies in critical illness in Blantyre which contributed significantly to the success of the work I present here. I owe particular thanks to Brian Faragher for insightful statistical review of chapters 3 and 4.

I was lucky enough to spend time as a visiting worker at the Wellcome Sanger Institute, and I thank all members of Team 216 for making me feel welcome, but particularly Mat Beale and Nick Thomson. Nick has been supportive of this fellowship from its inception and always given me time in a (very) busy schedule.

I thank my supervisors, Nick Feasey and Meilta Gordon for constant support, advice, and guidance.

Finally, I would like to thank my family, to whom this thesis is dedicated. The four years work that went into my fellowship involved sacrifices from all of them, and long periods of me being absent from home; but the fellowship could not have been a success without their love and support.



# Presentations and Publications

## Peer reviewed journal articles related to or arising from thesis

Lewis JM, Henrion, M, Rylance J. Validity of the qSOFA Score in Low- and Middle-Income Countries. *JAMA* 2018;320(19):2038-2039

Lewis JM, Feasey NA, Rylance J. Aetiology and outcomes of sepsis in adults in sub-Saharan Africa: a systematic review and meta-analysis. *Critical Care* 2019;23(1):212

Barr D\*, Lewis JM\*, Feasey NA, Andrews B, Kelly P, Lakhi S et al. Mycobacterium tuberculosis blood stream infection prevalence, diagnosis, and mortality hazard in HIV-positive adults: a systematic review and meta-analysis of individual patient data. *Lancet Infectious Diseases* (accepted).

\* Joint first author

Lewis JM, Abouyannis M, Katha G, Nyirenda M, Chatsika, G, Feasey NA, Rylance J. Population incidence and mortality of sepsis in an urban African setting 2013-2016. *Clinical Infectious Diseases* (online first 14 November 2019).

Lewis JM, Lester R, Garner P, Feasey NA. Gut mucosal colonisation with extended-spectrum beta-lactamase producing Enterobacteriaceae in sub-Saharan Africa: a systematic review and meta-analysis [version 1; peer review: 1 approved]. *Wellcome Open Research* 2019;4:160

Lewis JM, Lester R, Mphasa M, Banda R, Edwards T, Thomson NR, Feasey NA. Emergence of carbapenemase producing Enterobacteriaceae, Malawi. *Journal of Global Antimicrobial Resistance*. (under review).

## Oral presentations

Identifying drivers of ESBL-E carriage in Malawian adults: interim analysis from the DASSIM study. Malawi-Liverpool-Wellcome Clinical Research Programme Annual Scientific Meeting, Chester 2018

## Posters

Developing an antimicrobial strategy for sepsis in Malawi: the DASSIM study. Malawi-Liverpool-Wellcome Clinical Research Programme Annual Scientific Meeting, Malawi 2017



# Abbreviations

3TC: Lamivudine

AETC: Adult emergency and trauma centre

AMR: Antimicrobial resistance

aOR: Adjusted odds ratio

API: Analytical profile index

ARIBA: Antimicrobial resistance identification by assembly

ART: Antiretroviral therapy

AST: Antimicrobial sensitivity testing

BAPS: Bayesian analysis of population structure

BLAST: Basic local alignmeny search tool

BLI: Beta lactamase inhibitor

BP: Blood pressure

BPS: Best practice statement

BSAC: British Society for Antimicrobial Chemotherapy

BSI: Bloodstream infection

CCA: Complete case analysis

CCDEP: Center for Disease Dynamics, Economics and Policy

CD4: Cluster of differentiation 4

CI: Confidence interval

CoM: College of Medicine

COMREC: Malawi College of Medicine Research Ethics Committee

CPT: Cotrimoxazole preventative therapy

CRAG: Cryptococcal antigen

CSF: Cerebrospinal fluid

DAG: Directed acyclic graph

DASSIM: Developing an antimicrobial strategy for sepsis in Malawi

DBP: Diastolic blood pressure

DNA: Deoxyribonucleic acid

EDTA: Ethylenediaminetetraacetic acid

EFV: Efavirenz

EGDT: Early goal directed therapy

ELPD: Expected log predictive density

EPEC: Enteropathogenic *E. coli*

ESBL-E: Extended-spectrum beta-lactamase producing Enterobacteriaceae

ESBL: Extended-spectrum beta-lactam(ase)

ESICM: European Society of Intensive Care Medicine

ETEC: Enterotoxigenic *E. coli*

EU: European Union

FAMD: Factor analysis of mixed data

GCS: Glasgow coma scale

GDP: Gross domestic product

GEMS: Global enteric multicenter study

GLMM: Generalised linear mixed model

GNI: Gross national income

HDI: Human development index

HDU: High dependency unit

HIV: Human immunodeficiency virus

HR: Heart rate

HRQoL: Health related quality of life

IMAI: Integrated management of adolescent and adult illness

IPC: Infection prevention and control

IQR: Interquartile range

IS: Insertion sequence

ITU: Intensive therapy unit

IV: Intravenous

LA: Lumefantrine-artemether

LAM: lipoarabinomannan

LOESS: Locally estimated scatterplot smoothing

LSTM: Liverpool School of Tropical Medicine

LTFU: Lost to follow up

MAP: Mean arterial blood pressure

MAT: Microagglutination test

Mb: Megabases

MB: Megabytes

MCA: Multiple correspondence analysis

MCMC: Markov-chain Monte-Carlo

MDR: Multidrug resistant

MGE: Mobile genetic element

MGIT: Mycobacterial growth indicator tube

MLST: Multilocus sequence type

MLW: Malawi-Liverpool-Wellcome clinical research programme

mmHg: Millimetres of mercury

mRDT: Malaria rapid diagnostic test

MTB: *Mycobacterium tuberculosis*

MWK: Malawian Kwacha

NCBI: National Center for Biotechnology Information

NEQAS: National External Quality Assessment Service

NIHR: National Institute for Health Research

NUTS: No U-turn sampler

OR: Odds ratio

PaCO<sub>2</sub>: Partial pressure of carbon dioxide

PaO<sub>2</sub>: Partial pressure of oxygen

PCA: Principal component analysis

PCP: Pneumocystis jiroveci pneumonia

PCR: Polymerase chain reaction

PLR: Passive leg raise

QC: Quality control

QECH: Queen Elizabeth Central Hospital

qSOFA: Quick SOFA

QUAST: Quality assessment tool for genome assemblies

RCT: Randomised controlled trial

RDT: Rapid diagnostic test

RR: Respiratory rate

RR: Risk ratio/ respiratory rate

SBP: Systolic blood pressure

SC: Sequence cluster

SD: Standard deviation

SIRS: Systemic inflammatory response syndrome

SNP: Single nucleotide polymorphism

SOFA: Sequential organ failure assessment

SP = Sulfamethoxazole-pyrimethamine

SpO<sub>2</sub>: Capillary oxygen saturation

SRST: Short read sequence typing

sSA: Sub-Saharan Africa

ST: Sequence type

TB: Tuberculosis

TDF: Tenofovir

uLAM: Urinary lipoarabinomannan

USA: United States of America

WCC: White cell count

WGS: Whole genome sequencing

WHO: World health organisation

WSI: Wellcome Sanger institute

ZN: Ziehl–Neelsen



# Contents

<b>Collaborators and co-workers</b>	<b>5</b>
<b>Abstract</b>	<b>7</b>
<b>Acknowledgements</b>	<b>9</b>
<b>Presentations and Publications</b>	<b>11</b>
Peer reviewed journal articles related to or arising from thesis . . . . .	11
Oral presentations . . . . .	11
Posters . . . . .	12
<b>Abbreviations</b>	<b>13</b>
<b>1 Introduction</b>	<b>25</b>
1.1 Introduction . . . . .	25
1.2 Sepsis in sub-Saharan Africa . . . . .	26
1.2.1 Search strategy . . . . .	26
1.2.2 Statistical methods . . . . .	27
1.2.3 Defining sepsis . . . . .	27
1.2.4 Applicability of sepsis-3 definitions in sub-Saharan Africa . . . . .	28
1.2.5 Sepsis epidemiology in sub-Sahara Africa . . . . .	29
1.2.6 Sepsis aetiology in sub-Saharan Africa . . . . .	33
1.2.7 Sepsis management . . . . .	38
1.3 ESBL-E in sub-Saharan Africa . . . . .	44
1.3.1 Introduction: definition and classification of ESBL-E . . . . .	44
1.3.2 Global molecular epidemiology of ESBL-E: an overview . . . . .	46
1.3.3 Epidemiology of ESBL-E in sub-Saharan Africa . . . . .	51
1.4 Conclusions . . . . .	58
1.4.1 Specific aims . . . . .	59
1.5 Thesis overview . . . . .	59
1.6 Appendix . . . . .	60

1.6.1	Search terms for sepsis literature review . . . . .	60
1.6.2	Search terms for ESBL literature review . . . . .	60
<b>2</b>	<b>Methods</b>	<b>75</b>
2.1	Chapter Overview . . . . .	75
2.2	Study site . . . . .	75
2.2.1	Malawi . . . . .	75
2.2.2	Queen Elizabeth Central Hospital . . . . .	76
2.2.3	Participating Laboratories . . . . .	77
2.3	Clinical Study . . . . .	78
2.3.1	Objectives . . . . .	78
2.3.2	Recruitment criteria . . . . .	78
2.3.3	Study Visits and Patient Sampling . . . . .	79
2.3.4	Outcomes and sample size calculations . . . . .	81
2.4	Diagnostic Laboratory Procedures . . . . .	82
2.4.1	Point of care diagnostics . . . . .	82
2.4.2	Laboratory diagnostics . . . . .	82
2.4.3	Case definitions . . . . .	86
2.5	Molecular methods . . . . .	86
2.6	Statistical Analysis . . . . .	87
2.7	Study Team . . . . .	87
2.8	Data Collection and Storage . . . . .	87
2.9	Ethical Approval, Consent and Participant Remuneration . . . . .	88
<b>3</b>	<b>A clinical and microbiological description of sepsis in Blantyre, Malawi</b>	<b>91</b>
3.1	Chapter overview . . . . .	91
3.2	Introduction and chapter aims . . . . .	92
3.3	Methods . . . . .	92
3.4	Results . . . . .	94
3.4.1	Study population . . . . .	94
3.4.2	Baseline characteristics . . . . .	94
3.4.3	Admission physiology and laboratory investigations . . . . .	99
3.4.4	Aetiology . . . . .	101
3.4.5	Treatment . . . . .	104
3.4.6	Outcome . . . . .	107
3.4.7	Determinants of mortality . . . . .	109
3.5	Discussion . . . . .	114
3.5.1	Demographics and outcome: significant longer-term mortality . . . . .	114
3.5.2	Aetiology: TB dominates as a cause of sepsis . . . . .	116
3.5.3	Determinants of mortality . . . . .	116



3.5.4	Limitations . . . . .	119
3.6	Conclusions and further work . . . . .	120
<b>4</b>	<b>Modelling to identify determinants of sepsis mortality</b>	<b>121</b>
4.1	Chapter overview . . . . .	121
4.2	Introduction and chapter aims . . . . .	121
4.3	Methods . . . . .	124
4.4	Results . . . . .	127
4.4.1	Exploring time-to antibacterials and IV fluid as determinants of mortality . . . . .	130
4.4.2	Propensity score matching and subgroup analysis . . . . .	132
4.5	Discussion . . . . .	134
4.5.1	Limitations . . . . .	137
4.6	Conclusions and further work . . . . .	138
4.7	Appendix . . . . .	138
<b>5</b>	<b>ESBL-E carriage in Malawian adults in health and disease</b>	<b>143</b>
5.1	Chapter Overview . . . . .	143
5.2	Introduction and chapter aims . . . . .	144
5.3	Methods . . . . .	144
5.4	Results . . . . .	146
5.4.1	Study population . . . . .	146
5.4.2	Exposures during the study period . . . . .	153
5.4.3	ESBL-E colonisation . . . . .	156
5.4.4	Associations of ESBL colonisation . . . . .	156
5.5	Discussion . . . . .	163
5.5.1	Limitations . . . . .	165
5.6	Conclusions and further work . . . . .	165
<b>6</b>	<b>The genomic landscape of ESBL producing <i>E. coli</i> in Blantyre, Malawi</b>	<b>167</b>
6.1	Chapter overview . . . . .	167
6.2	Introduction and chapter aims . . . . .	168
6.3	Methods . . . . .	168
6.3.1	Bioinformatic pipeline . . . . .	168
6.3.2	Global <i>E. coli</i> collection . . . . .	169
6.3.3	Statistical analysis . . . . .	170
6.4	Results . . . . .	170
6.4.1	Samples and quality assurance and control . . . . .	170
6.4.2	Phylogroup, MLST and core genome phylogeny of study isolates . . . . .	171
6.4.3	Study isolates in a global context . . . . .	173

6.4.4	Antimicrobial resistance determinants . . . . .	178
6.4.5	Plasmid replicons . . . . .	185
6.5	Discussion . . . . .	186
6.5.1	Genomic landscape of ESBL <i>E. coli</i> in Malawi: global diversity and high-risk clones . . . . .	186
6.5.2	Antimicrobial resistance determinants: domination of <i>bla</i> <sub>CTXM-15</sub> and emergence of carbapenemases . . . . .	190
6.5.3	Study limitations . . . . .	191
6.5.4	Conclusions and further work . . . . .	192
6.6	Appendix . . . . .	193
<b>7</b>	<b>Whole genome sequencing to track longitudinal ESBL-E colonisation</b>	<b>215</b>
7.1	Chapter overview . . . . .	215
7.2	Introduction and chapter aims . . . . .	216
7.3	Methods . . . . .	216
7.4	Results . . . . .	218
7.4.1	Hierarchical BAPS clustering of core gene pseudosequences . . . . .	218
7.4.2	ESBL-clusters . . . . .	219
7.4.3	Assessing for healthcare-associated lineages . . . . .	221
7.4.4	Assessing for within-patient conservation of lineage or MGE . . . . .	221
7.5	Discussion . . . . .	224
7.5.1	Limitations . . . . .	227
7.6	Conclusions and further work . . . . .	228
<b>8</b>	<b>Longitudinal Markov models of ESBL-E carriage</b>	<b>235</b>
8.1	Chapter Overview . . . . .	235
8.2	Introduction and chapter aims . . . . .	235
8.3	Methods . . . . .	237
8.3.1	Developing the models used in this chapter . . . . .	237
8.3.2	General form of likelihood . . . . .	237
8.3.3	Markov model likelihood . . . . .	238
8.3.4	Incorporating covariates: a proportional hazard model . . . . .	239
8.3.5	Building and fitting models . . . . .	241
8.3.6	Assessing goodness of fit . . . . .	242
8.3.7	Exploring differences in carriage dynamics by bacterial species and <i>E. coli</i> genotype . . . . .	243
8.3.8	Simulations from the posterior . . . . .	244
8.4	Results . . . . .	244
8.4.1	The effect of antibacterials and hospitalisation on ESBL-E carriage .	244
8.4.2	Exploring bacterial species and genotype differences in carriage dynamics	248

8.4.3	Simulation of different antibacterial and hospitalisation scenarios . .	250
8.5	Discussion . . . . .	250
8.5.1	Limitations . . . . .	254
8.6	Conclusion and further work . . . . .	256
8.7	Appendix . . . . .	257
<b>9</b>	<b>Conclusions and further work</b>	<b>271</b>
9.1	Introduction . . . . .	271
9.2	Summary of findings . . . . .	271
9.3	Conclusions and future research priorities . . . . .	273
	<b>Appendix: R Session Information</b>	<b>277</b>
	<b>References</b>	<b>281</b>



# List of Figures

1.1	Pooled sepsis inpatient mortality . . . . .	34
1.2	3GC resistance worldwide . . . . .	48
1.3	Invasive ESBL-E in <i>E. coli</i> and <i>K. pneumoniae</i> in sSA . . . . .	53
1.4	Invasive ESBL-E in <i>E. coli</i> and <i>K. pneumoniae</i> in sSA BSI . . . . .	54
1.5	ESBL enzymes in invasive ESBL-E in sSA. . . . .	55
1.6	Prevalence of ESBL-E gut mucosal carriage in sSA . . . . .	56
1.7	ESBL genes in gut carriage in sSA . . . . .	57
2.1	Map of study site . . . . .	76
2.2	Overview of patient sampling schedule . . . . .	81
2.3	Urinary LAM refrence card . . . . .	85
2.4	Representative example of chromogenic agar appearance. . . . .	85
2.5	Example of combination disc ESBL confirmatory testing. . . . .	86
3.1	Study recruitment and follow up. . . . .	95
3.2	Distribution of reported time of ART (months) . . . . .	98
3.3	Symptoms of recruited participants . . . . .	99
3.4	Venn diagram of positive diagnostic tests . . . . .	104
3.5	UpSet plot of positive diagnostic tests . . . . .	105
3.6	Pathogenic isolates recovered from aerobic blood culture. . . . .	105
3.7	Antimicrobial and fluid administration as a function of time. . . . .	106
3.8	Kaplan-Meier survival curves. . . . .	108
3.9	Health-related QoL following admission with sepsis. . . . .	108
4.1	Hypothesised causal structre of mortality in sepsis . . . . .	125
4.2	Composition of FAMD variables . . . . .	129
4.3	Association of FAMD dimensions with mortality . . . . .	130
4.4	Modelling effects of treatments on death in sepsis following dimensionality reduction. . . . .	131
4.5	Modelling associations of time to antimicrobials and death by 28 days . . .	133

4.6	Propensity-score matched and subgroup analysis of effect of TB therapy on mortality. . . . .	135
4.7	Variable distributions in sepsis cohort following propensity score matching. . . . .	142
5.1	ESBL carriage study recruitment and follow up. . . . .	148
5.2	Missing samples and variation in sample collection time . . . . .	149
5.3	Hospital and antibacterial exposure of participants . . . . .	154
5.4	ESBL colonisation prevalence as a function of time . . . . .	157
5.5	Heatmap showing distribution of ESBL positive samples . . . . .	158
5.6	Species and antimicrobial sensitivities of cultured ESBL-E . . . . .	159
5.7	Univariable and multivariable associations of ESBL-E acquisition by 28 days . . . . .	162
6.1	Species read assignment of sequenced isolates . . . . .	172
6.2	N50 and total assembly length for included assemblies . . . . .	173
6.3	<i>E. coli</i> multilocus sequence type distribution . . . . .	174
6.4	Phylogenetic tree of study <i>E. coli</i> . . . . .	175
6.5	Phylogenetic tree of study <i>E. coli</i> in global context . . . . .	176
6.6	Phylogeny of ST131, ST410 and ST167 <i>E. coli</i> . . . . .	177
6.7	Frequency distribution of AMR genes . . . . .	180
6.8	Quinolone resistance mutations and phenotypic resistance . . . . .	181
6.9	Aminoglycoside resistance mutations and phenotypic resistance . . . . .	183
6.10	Chloramphenicol and cotrimoxazole resistance mutations and phenotypic resistance . . . . .	185
6.11	Co-occurrence and lineage association of AMR genes . . . . .	187
6.12	Networks of co-occurring AMR genes . . . . .	188
6.13	AMR genes and plasmid replicons mapped back to tree . . . . .	188
7.1	Core gene hierBAPS clusters . . . . .	220
7.2	ESBL-clusters mapped back to phylogeny . . . . .	222
7.3	Healthcare association of isolates mapped to phylogeny . . . . .	223
7.4	Within- and between pairwise SNP distance of isolates . . . . .	224
7.5	Within-patient ESBL-cluster and hierBAPS clustering as a function of time . . . . .	225
7.6	Summary statistics for ESBL-contig clusters . . . . .	229
7.7	AMR genes, IS and plasmid replicon content of ESBL-clusters, <i>blaCTXM15</i> . . . . .	230
7.8	AMR genes, IS and plasmid replicon content of ESBL-clusters, <i>blaCTXM27</i> and <i>blaSHV12</i> . . . . .	231
7.9	AMR genes, IS and plasmid replicon content of ESBL-clusters: <i>blaCTXM14</i> and <i>blaCTXM9</i> . . . . .	232
7.10	AMR genes, IS and plasmid replicon content of ESBL-clusters: <i>blaCTXM16</i> , <i>blaCTXM3</i> and <i>blaCTXM3</i> . . . . .	233

8.1	Two state ESBL-E model . . . . .	240
8.2	Parameter estimates from Markov models . . . . .	246
8.3	Predicted proportion of ESBL-E positive samples, stratified by arm. . . . .	247
8.4	Parameter estimates from models of bacterial species and genotype carriage	249
8.5	Simulations of different antibacterial and hospitalisation scenarios . . . . .	251
8.6	ESBL-E carriage model pairwise posterior parameter estimates . . . . .	258





# List of Tables

1.1	Search terms for fever studies . . . . .	27
1.2	Characteristics of patients recruited to sSA sepsis studies . . . . .	31
1.3	Aetiology of sepsis in sSA . . . . .	35
1.4	BSI isolates in sepsis in sSA . . . . .	35
1.5	Surviving sepsis campaign guidelines . . . . .	39
1.6	ESICM low resource setting sepsis recommendations . . . . .	40
1.7	ESBL classification. Adapted from [139] . . . . .	45
1.8	Sepsis diagnostic criteria . . . . .	61
1.9	Sequential organ failure assessment (SOFA) score . . . . .	62
1.10	Selected causes of fever in sSA since 2013 . . . . .	63
1.11	included studies providing an estimate of proportion of ESBL producers in invasive <i>E. coli</i> and <i>K. pneumoniae</i> isolates in sSA. . . . .	70
1.12	Included studies providing estimate of prevalence of ESBL-E gut mucosal colonisation in sSA. . . . .	72
2.1	Study inclusion criteria . . . . .	79
3.1	Participant characteristics . . . . .	96
3.2	Prehospital healthcare seeking and antimicrobial exposure . . . . .	98
3.3	Admission physiology, haematology and biochemistry . . . . .	100
3.4	Number of diagnoses . . . . .	102
3.5	Diagnostic tests performed and results, stratified by HIV status. . . . .	103
3.6	Door-to-antimicrobial times. . . . .	107
3.7	Day 28, 90 and 180 mortality stratified by HIV status . . . . .	109
3.8	Bivariate associations with death by 28 days . . . . .	110
3.9	Unadjusted and adjusted odds ratios of death by 28 days . . . . .	113
4.1	Unadjusted and adjusted odds ratios of death by 28 days in sepsis from Bayesian logistic regression, for complete case analysis (CCA) and following multiple imputation of missing data. . . . .	128
4.2	Adjusted odds ratio of death by 28 days per hour delay in antibacterials . .	132

4.3	Bivariable associations of receipt of TB treatment in sepsis . . . . .	139
5.1	Participant characteristics . . . . .	150
5.2	Antimicrobial and hospital exposure stratified by arm . . . . .	155
5.3	ESBL carriage stratified by arm and visit . . . . .	156
5.4	Univariable and multivariable associations of ESBL colonisation at enrolment	161
6.1	Phylogroup distribution of sequenced <i>E. coli</i> isolates. . . . .	173
6.2	Frequency of identification of plasmid replicons . . . . .	185
6.3	Details of samples included in global phylogeny . . . . .	193
8.1	Estimates (and standard error) of pairwise expected log predictive density (ELPD) differences for all models . . . . .	248
8.2	Parameter estimates (and 95% confidence intervals) from model 2 . . . . .	250

# Chapter 1

## Introduction

### 1.1 Introduction

The syndrome of sepsis has been described since antiquity; from Hippocrates to Galen and Semmelweis, the potentially serious systemic consequences of a localised infection have long been recognized. The word sepsis arises from the Greek  $\sigma\eta\psi\iota\varsigma$  meaning decomposition and was described by Hippocrates as a dangerous putrefaction in the body[1]. Modern definitions of sepsis conceptualise it as a syndrome of life threatening organ dysfunction due to a dysregulated host response to infection[2], but despite increased understanding of its pathogenesis[3], mortality from sepsis remains high. Progress has been made in improving sepsis mortality in high income settings[4,5], through timely application of basic care[6,7]: early appropriate antimicrobials, aggressive fluid resuscitation and organ support largely in a critical care environment. Limited data from low resource settings including sub-Saharan Africa (sSA) suggest that mortality remains high[8], and increasing evidence suggests that exporting high-income setting sepsis protocols to sSA has the potential to do harm[9,10]. Data to guide sepsis management protocols for sSA are urgently needed.

Data on sepsis aetiology from sSA to guide antimicrobial strategies are lacking; currently, in Blantyre Malawi, for example, empirical management of sepsis is the norm and patients often receive prolonged empiric courses of broad spectrum antimicrobials – largely ceftriaxone, a third-generation cephalosporin antibiotic. Limited data suggest that tuberculosis, arboviruses and bacterial zoonoses may be important causes of severe febrile illness in sSA[11–14], pathogens which largely go untreated by ceftriaxone. On a population level invasive *Escherichia coli* and *Klebsiella pneumoniae* bacteria are showing an alarming increase in ceftriaxone resistance since the drug was introduced in Malawi in 2005[15,16]. The majority of these resistant bacteria are so-called extended-spectrum  $\beta$ -lactamase producers (ESBL-producers) and are often untreatable with locally available antimicrobials. Novel antimicrobial strategies are needed to safely preserve ceftriaxone - often a first and last line

antibiotic in Malawi - for those who need it.

It is the hypothesis of this thesis, then, that sepsis in Malawi is caused by a wide variety of infections that are currently unrecognised and untreated, and that this is contributing to high sepsis mortality. Conversely, prolonged ceftriaxone exposure in sepsis survivors is causing acquisition and carriage of resistant bacteria (principally ESBL Enterobacteriaceae, henceforth ESBL-E) and their transportation into the community. I ask the question as to whether novel antimicrobial strategies in sSA can not only improve outcomes in sepsis, but can minimise pressure for antimicrobial resistance (AMR) by altering the way in which we use antibacterials in these very sick patients. Before addressing this question I will review, in this chapter, the definitions, epidemiology, aetiology and management of sepsis, with a focus on aetiology and antimicrobial treatment followed by the epidemiology and drivers of ESBL-E carriage, both with a focus on sSA.

## 1.2 Sepsis in sub-Saharan Africa

### 1.2.1 Search strategy

A review of the literature was undertaken to identify prospective cohort, case control studies or randomised controlled trials (RCTs) of sepsis in sub-Saharan Africa with the search terms shown in the appendix to this chapter. Pubmed and Scopus were searched, yielding 5460 unique studies on 17 July 2018. Inclusion criteria were any prospective cohort, RCT or case-control studies of sepsis in sSA (defined as taking place in the countries listed in search terms panel) recruiting patients using sepsis 1,2 or 3 definitions. Abstract review was undertaken resulting in inclusion of 91 studies for full text review. Eleven publications providing data on eight prospective cohorts[17–24] and three intervention studies (two RCTs[25,26] and one before-after intervention[27]) were identified. These data inform the following review, alongside non-systematically searched studies examining sepsis in high-resource settings.

In order to put sepsis aetiology data in context, systematic searches of the Pubmed and Scopus databases for leptospirosis, brucellosis, Q fever, rickettsioses, arboviruses (dengue, or chikungunya) and histoplasmosis prevalence in unselected sepsis or fever cohorts in sSA were undertaken. Because a recent systematic review[12] has examined these pathogens up to 2013 (see “sepsis aetiology” below), the date of these searches were restricted the 2014 to the present. Any studies examining disease prevalence in cohorts of febrile adults or children were included; outbreaks were excluded. Studies where the inclusion criteria were not clear (including those with, for example, “suspected leptospirosis” with no further details) were excluded. Finally, systematic searches of *Pneumocystis jiroveci* pneumonia (PCP) were made using the search terms below; because a recent systematic review has

examined the role of PCP up to 2015, the date on this search was restricted to 2015 or later. Table 1.1 shows the search terms, number of hits and number of included studies after full text review: nine studies provided data on Leptospirosis[28–36], seven on Brucellosis[37–43], seven on Q-fever[35,39,44–47], six on Rickettsioses[35,44,48–51], eighteen on Dengue[29,31,35,36,44,50,52–63], thirteen on Chikungunya[31,36,50,53,56,58,60–66], three on Zika [59–61], two on Histoplasmosis[67,68] and none on PCP. Details of the included studies are provided below.

Table 1.1: Search terms for fever studies

Organism	Search	n abstracts	n included
Leptospirosis	Leptospir AND	187	9
Brucellosis	Brucell AND	123	7
Q-fever	((Q fever) OR (coxiella)) AND	315	7
Rickettsioses	(Ricketts OR typhus OR (spotted fever)) AND	375	6
Arboviruses	(dengue OR chikungunya OR arbovir) AND	1422	Dengue 18, Chikungunya 13, Zika 3
Histoplasmosis	Histoplasm AND	72	2
PCP	(((PCP) OR pneumocystis) OR "pneumocystis carini*") OR "pneumocystis jiroveci")) AND	87	0

*Note:*

All searches included the sSA country list in addition to the disease-specific terms above.

### 1.2.2 Statistical methods

Largely, narrative review of identified sepsis cohorts was undertaken, but meta analysis was used to summarise outcomes. 28 or 30-day sepsis mortality were extracted from identified studies or, if 28- or 30-day data were not available, in-hospital mortality was used. For interventional studies, in order to reflect the “usual-care” mortality, only the usual care arms were included. Pooled mortality estimates were then generated using a random effect meta-analysis of proportions with a generalised linear mixed model (GLMM, the so called binomial-normal model) using the *meta* package v.4.9.5[69] in R v3.6.0. Exact binomial 95% confidence intervals for all proportions were used throughout.

### 1.2.3 Defining sepsis

Sepsis is a heterogeneous syndrome, with no diagnostic gold standard. In 1991 the first modern sepsis diagnostic criteria were defined in a consensus conference of key opinion makers[70] (Table 1.8, chapter appendix). Sepsis was defined as the presence of the systemic

inflammatory response syndrome (SIRS) plus infection, with a gradient of severity increasing through severe sepsis (sepsis plus organ dysfunction) to septic shock. These definitions were widely adopted as entry points into clinical trials, but ongoing concerns that SIRS was both insensitive and non-specific for the diagnosis of sepsis led to an expansion of the diagnostic criteria in 2001[71] again by expert consensus. Despite these revised guidelines the SIRS criteria largely continued to be preferred both as the entry point to clinical trials of sepsis and in clinical practice until the development of the current sepsis-3 definitions in 2016[2].

The sepsis-3 definitions redefined sepsis as “life threatening organ dysfunction triggered by infection”, a definition that rendered the sepsis-2 severe sepsis category obsolete. In contrast to the previous diagnostic criteria that had relied largely on expert opinion, the sepsis-3 criteria attempted to use a probabilistic approach to defining sepsis, by mandating that sepsis should be associated with excess mortality. The sequential organ dysfunction score (SOFA, Table 1.9, chapter appendix), an organ-dysfunction score already in use in high income settings, and shown to be associated with mortality[72] was selected to operationalise the definition of sepsis. An acute change in SOFA of 2 or more points defines sepsis under sepsis-3.

Mindful that the SOFA score requires a large number of variables and is difficult to apply at the bedside, the consensus guideline group suggest the use of a simpler score, quick SOFA to identify patients who may have sepsis. Any two of: altered mental status, SBP < 100mmHg or respiratory rate > 22 breaths min<sup>-1</sup> defines a positive qSOFA score. qSOFA does not define sepsis; rather, under sepsis-3 patients with a qSOFA score of 2 or more are at increased risk of poor outcomes and should be screened for sepsis using a full SOFA score. The qSOFA was derived by identifying factors associated with mortality in large datasets of patients with infection from the United States and validated in further US and German datasets; in these datasets it showed good discriminant ability to predict mortality, equivalent to full SOFA score outside the intensive therapy unit (ITU)[73].

Finally, sepsis-3 defines septic shock as persistent hypotension requiring vasopressors to maintain mean arterial blood pressure (MAP) above 65mmHg and serum lactate greater than 2mmol L<sup>-1</sup>. This definition was arrived at by a combination of consensus and systematic review to identify potential defining variables and validation in large datasets from the United States, where it was found to be strongly associated with mortality[74].

#### 1.2.4 Applicability of sepsis-3 definitions in sub-Saharan Africa

Application of the sepsis-3 definitions, both in terms of clinical use and as inclusion criteria for research studies in sub-Saharan African low resource settings, is problematic. Several of the domains of SOFA require the results of blood tests, which may not be available. In Blantyre, and elsewhere in sSA, intensive organ support with inotropes or mechanical

ventilation (invasive or non-invasive) may not be available[75] or be difficult to access[76], yet use of these treatment modalities form components of the SOFA score. Both lactate measurement and inotropic support may be unavailable in some settings and yet these define septic shock. Five studies have validated the qSOFA score in sub-Saharan African settings[22,77–80] and found variable discriminant ability for mortality but it is not clear how this score should be deployed in this setting; no studies have been undertaken to link qSOFA score to clinical action, and it is not intended to define sepsis under sepsis-3. The optimal sepsis definitions (both clinical and for research) for sSA are therefore not clear.

### 1.2.5 Sepsis epidemiology in sub-Sahara Africa

#### 1.2.5.1 Incidence

The changing case definition of sepsis over time hampers estimation of incidence even in high-income settings, furthermore sepsis is not included in global burden of disease estimates. Different methods of defining sepsis from disease registries can result in very different estimates[81], but a recent systematic review and meta-analysis of 27 studies from 9 high income countries found a recent population incidence rate of 437/100,000 person-years (95% CI 334-571) for sepsis and 270 (95% CI 176 – 412) for severe sepsis with an increasing incidence over time from 1979 to 2015[4]. Crudely extrapolating these estimates to the worldwide population would result in 20.7 million sepsis and 10.7 million severe sepsis cases a year, largely in low resource settings. However, no data are available from low or middle income settings and these estimates must be treated with caution.

#### 1.2.5.2 Risk factors: the sepsis population in sub-Saharan Africa

In high-income settings, risk factors for sepsis have been identified, though once again changing definitions as well as a lack of large scale community based studies make it difficult to draw definitive conclusions. However, chronic diseases (including HIV) and immunosuppressive agents have been associated with increased sepsis incidence, as well as older age[82,83]. In the United States, male sex and black ethnicity (vs white) and poverty are associated with increased sepsis incidence and severity[84].

Though equivalent studies aiming to identify risk factors for sepsis in adults in sSA are lacking, it is clear from the available data that HIV-infection is the dominant risk factor there. Summary patient demographics from the 10 identified sepsis studies are shown in Table 1.2; of 2788 included patients with available HIV status, 69% (1809/2788) were HIV infected, and often with advanced disease; of 1278 HIV-infected patients from 5 studies with available CD4 count the median CD4 count ranges from 52-168 cells/ $\mu$ L. In keeping with the epidemiology of the HIV epidemic in Africa, these patients are young, with average ages

(variably reported as mean or median) ranging from 30-39 across the studies. These studies recruited an equal proportion of males and females (1444/2812 males, 51%), suggesting that sex is not a risk factor.

These data contrast sharply with the sepsis population in high income settings, from whom the majority of sepsis data have been generated, and who are older and mostly HIV uninfected[7,82,85]. The need for data from sSA to guide sepsis treatment protocols, rather than extrapolating from the high-income setting sepsis population, is clear.



Table 1.2: Characteristics of patients recruited to sSA sepsis studies

Study	Type	Year	Country	Inc. criteria	n	Male	Age	HIV infected	Median CD4
Jacob 2009	Cohort	2006	Uganda	Severe sepsis	382	156/382 (41%)	34.8 (11.2)	320/382 (85%)	52 (16-131)
Jacob 2012 <sup>a</sup>	Before-after	2006	Uganda	Severe sepsis	245	95/245 (39%)	34 (28-41)	207/245 (86%)	43 (11-178)
		2008-09			426	207/426 (49%)	34 (27-40)	362/426 (85%)	63 (15-178)
Waitt 2015	Cohort	2008-09	Malawi	Sepsis	213	87/213 (41%)	30 (25-39)	161/213 (76%)	NR
Ssekitoleko 2011 (1) <sup>b</sup>	Cohort	2009	Uganda	Sepsis	96 <sup>b</sup>	193/418 <sup>b</sup> (46%)	35.1 (12.0)	331/418 <sup>b</sup> (83%)	NR
Ssekitoleko 2011 (2)	Cohort	2009	Uganda	Sepsis	150	94/150 (63%)	35 (13)	96/150 (64%)	NR
Chimese 2012	Cohort	2010	Zambia	Sepsis	161	79/161 (49%)	39 (15.6)	110/138 (80%)	NR
Andrews 2014	RCT	2012	Zambia	Severe sepsis	112	58/109 (53%)	35 (1.4)	88/109 (81%)	NR
Auma 2013	Cohort	2012	Uganda	Sepsis	216	106/216 (49%)	32 (27-43)	122/216 (56%)	NR
Andrews 2017	RCT	2012-13	Zambia	Severe sepsis	209	117/209 (56%)	36.7 (12.4)	187/209 (89.5%)	66 (21-143)
Huson 2014	Cohort	2012-13	Gabon	Sepsis	384	142/382 (37%)	34 (25-46)	77/384 (20%)	168 (61-438)
Amir 2016	Cohort	2014-15	Uganda	Severe sepsis	218	110/218 (50%)	35 (26-50)	125/218 (57%)	78 (20-202)

*Note:*

RCT = randomised controlled trial. All studies use a modified sepsis-2 definition of sepsis or severe sepsis. Age is given as median (IQR) or mean (SD). Units of CD4 count are cells/microlitre.

<sup>a</sup> Jacob 2012 includes two cohorts of patients – results shown for both separately.

<sup>b</sup> Ssekitoleko 2011 (1) includes some participants also included in Jacob 2009; the n here (n=96) includes only participants not included in Jacob 2009 but the proportion of male and HIV-infected participants includes all patients including those from Jacob 2009 (n= 418) as they are not disaggregated in the publication.

### 1.2.5.3 Outcomes

Summary 28/30 day mortality outcomes for sepsis and severe sepsis in sSA from the identified studies are presented in Figure 1.1. It is clear that there is significant heterogeneity in outcomes of sepsis and severe sepsis in sSA, likely reflecting diverse patient and pathogen populations and variation in availability of available resources. This heterogeneity means that summary estimates should be interpreted with extreme caution but severe sepsis (49% [95% CI 39-58]), as expected, seems to carry a higher mortality hazard than sepsis (23% [95% CI 12-38]). Data of outcomes beyond 30 days are absent.

How does this compare to high income settings? A recent meta-analysis of population level estimates in high income settings found that a pooled sepsis 30-day mortality estimate of 17% (95% CI 11-26%)[4], though even older cohort studies as well as the more recent large sepsis-3 derivation cohorts have found considerably lower mortalities for sepsis (as defined by sepsis-2) ranging from 4-7%[73,86,87]. Most recent (largely post-2005) estimates of 30-day mortality from severe sepsis range from 18-29%[4,5,81,85,88]. It seems likely therefore, that both sepsis and severe sepsis 30-day mortality is considerably higher in sSA than in high-income settings.

In the longer term, sepsis mortality continues to rise after the usual sepsis-study primary end point of 28 or 30 days, though data from sSA are absent. A systematic review in 2010 of long term sepsis mortality identified 26 studies (with none from low-resource settings) that reported long term sepsis mortality; 1 year mortality ranged from 22-72%, increasing to 45-75% at greater than 3 years[89]. Both short and long term morbidity is formidable also, though, once again, data from low income settings including sSA are absent[90–94] and health-related quality of life in sepsis survivors in high-income settings have been found to be persistently below population norms[89]. Long term sepsis outcomes in sSA are unknown.

### 1.2.6 Sepsis aetiology in sub-Saharan Africa

The 11 identified prospective sepsis studies in sSA carried out various combinations of diagnostic testing for malaria (either microscopy or rapid diagnostic test) and aerobic and mycobacterial blood culture; a summary is shown in Table 1.3 and 1.4. The commonest bloodstream infection (BSI) in all studies where mycobacterial blood cultures were carried out was tuberculosis – present in a higher proportion than of all BSI isolates from aerobic culture combined - though it is important to note that mycobacterial blood cultures in most studies were carried out in HIV infected people and bacteraemic tuberculosis was almost exclusively HIV-associated. With the exception of one study, malaria was less common than BSI, highlighting the importance of non-malarial fever in sSA as malaria control efforts reduce the burden of malaria.

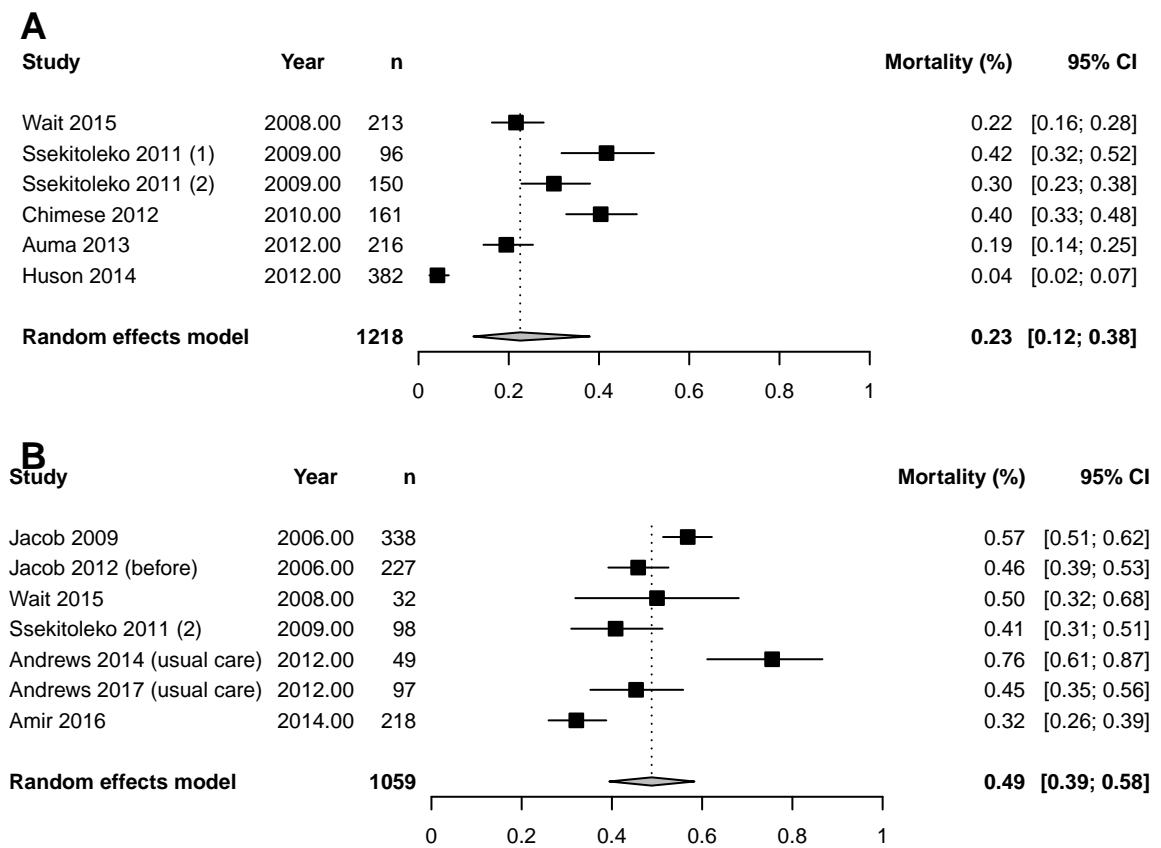


Figure 1.1: Pooled sepsis (A, top) and severe sepsis (B, bottom) inpatient mortality in sSA

Table 1.3: Aetiology of sepsis in sSA

Study	BSI	MTB BSI	Malaria
Jacob 2009	48/382 (13%)	156/382 (22%)	34.8 (15%)
Jacob 2012	83/671 (12%)	104/576 (18%)	83/671 (12%)
Waitt 2015	33/213 (15%)	ND	26/213 (12%)
Ssekitoleko 2011 (1)	ND	ND	ND
Ssekitoleko 2011 (2)	39/150 (26%)	ND	7/150 (5%)
Chimese 2012	27/161 (17%)	ND	ND
Andrews 2014	26/109 (24%)	32/81 (40%)	2/109 (2%)
Auma 2013	41/216 (19%)	ND	9/216 (4%)
Andrews 2017	29/209 (14%)	43/187 (23%)	3/47 (6%)
Huson 2014	39/384 (10%)	NR	130/384 (33%)
Amir 2016	ND	ND	ND
<b>TOTAL</b>	<b>365/2493 (15%)</b>	<b>234/1093 (21%)</b>	<b>311/2139 (15%)</b>

Table 1.4: BSI isolates in sepsis in sSA

Organism	N
<i>S. aureus</i>	109
Non-Typhoidal Salmonellae	84
<i>S. pneumoniae</i>	67
Non-salmonella Enterobacteriaceae	46
<i>Cryptococcus</i> spp.	20
<i>S. Typhi</i>	6
Other	33
<b>TOTAL</b>	<b>365</b>

*Note:*

Excluded are coagulase-negative Staphylococci, alpha-haemolytic Streptococci other than *Pneumococcus*, *Bacillus* spp. and Micrococci as likely contaminants.

### 1.2.6.1 Tuberculosis

Beyond the studies of sepsis described above, there is ample evidence that disseminated tuberculosis is a significant cause of hospitalisation and death in people living with HIV in sSA, but optimum management in the context of critical illness is less clear. Mycobacterial blood culture is a difficult diagnostic tool to use in clinical practice - it requires laboratory infrastructure and usually takes many weeks to become positive - but when carried out hospitalised adults in sSA the prevalence of *Mycobacterium tuberculosis* bloodstream infection (MTB BSI) is between 2.5-23%, and almost universally restricted to participants with HIV[14,95–108]. The recent STAMP study in Malawi and South Africa found a mortality benefit in some prespecified subgroups of a strategy of screening all HIV-infected inpatients for TB using urinary lipoarabinomannan[109]. Autopsy studies have persistently found evidence of TB in a significant proportion of HIV-infected people who die in hospital - 43% in one meta analysis - which is often missed ante mortem, and is very often disseminated[110].

Both of these findings strongly suggest a high burden of undiagnosed disseminated TB in HIV-infected inpatients. The WHO, recognising this, has published guidelines on the management of smear negative tuberculosis in the seriously unwell[111] which suggest a trial of broad-spectrum antimicrobials for 3-5 days and prompt initiation of TB therapy if there is no response. Though based on expert opinion, these guidelines have been shown to improve outcomes in South Africa[112]. It is unknown whether delaying TB therapy in this way is associated with higher mortality in the critically unwell, analogously to antibacterial delay in sepsis in high income settings.

#### 1.2.6.2 Bacterial zoonoses, rickettsioses and arboviruses

There are several reasons to suspect that aetiologic agents other than bacterial BSI and tuberculosis may be significant in sSA, though data in sepsis are sparse. Studies of febrile illness in sSA have implicated rickettsioses, arboviruses and bacterial zoonoses as causes of fever, accounting for a third of fever in hospitalised adults in one study in Tanzania[14]. Historically, however, data on these pathogens have been lacking. A 2015 systematic review of fever aetiology in LMIC (considering studies from 1980-2013) found that small numbers of patients had been systematically screened for these pathogens: in sSA 40/453 (8.8%) of adults with fever fulfilled diagnostic criteria for Leptospirosis, 16/453 (3.5%) for Brucellosis, 36/450 (8.0%) for spotted fever group Rickettsiosis, 24/482 (5%) for Q-fever and 55/700 (7.9%) for Chikungunya[12].

Increasing interest in non-malarial fever, however, has meant that data are accumulating from different settings in sSA, post-2013, as identified by the systematic review of the literature performed for this thesis. Details of the studies identified from this review are shown in Table 1.10 in the chapter appendix. These data highlight, firstly, the heterogeneity in diagnostics which are used for these pathogens – a combination of serology, PCR and antigen testing (often not using gold-standard case definitions), and secondly, the spatial and temporal heterogeneity across the continent.

These studies also demonstrate an increase, post-2013, in the use of molecular tests, particularly multiplex PCR assays (TaqMan array cards or PCR macroarrays) to detect multiple pathogens in fever aetiology studies. Despite the attractiveness of these assays – the ability to detect tens of pathogens in one assay on one body fluid sample – many infections will have only transiently detectable pathogen genetic material in blood and as such may have limited sensitivity. The post-2013 fever aetiology data strongly suggest paired sera will maximise the diagnostic yield of bacterial zoonoses and rickettsioses: for example, in studies of leptospirosis using PCR only 23/2533 (0.9%) of samples were positive versus 75/1464 (5.1%) in studies using paired sera; for Q-fever 9/3811 (0.2%) of samples were positive in PCR only studies versus 25/370 (6.8%) for paired sera studies; for brucellosis

PCR only studies 15/1005 (1.5%) of samples were positive versus 39/562 (6.9%) for paired sera studies; and for rickettsioses 55/1932 (2.8%) of samples were positive for PCR studies vs paired sera 63/364 (17%). Some care must be taken with this conclusion: there are no studies that aim to directly compare paired sera and PCR assays for diagnosis of febrile illness, so the possibility of confounding remains.

Available data therefore suggest that bacterial zoonoses, rickettsioses and arboviruses are significant causes of febrile illness in sSA. Their role in sepsis however is unknown. Only two studies have directly addressed the question of sepsis aetiology beyond BSI, malaria and TB: the first[11] performed PCR for 43 pathogens (using a TaqMan array card) including viruses (including dengue, chikungunya, and causes of viral haemorrhagic fever), bacteria (including *S. pneumoniae*, *E. coli*, *Salmonella* spp., *S. aureus* as well as *Coxiella burnetti*, *Rickettsia* spp., *Brucella* spp. and *Leptospira* spp.), Mycobacterial (including *M. tuberculosis* (MTB) and *M. avium* complex), fungal (*Cryptococcus* and *Histoplasma* spp.) and parasitic (including malaria) on a convenience sample of 336 stored plasma samples from a Ugandan sepsis study. In keeping with the original study, MTB was frequently identified as was pneumococcus and malaria. Cytomegalovirus (CMV) was detected in 139/336 (41%) of patients, and was found to be independently associated with death, a finding which has been seen in sepsis studies in high-income settings[113] and may be related to the immune paresis of sepsis and CMV viraemia rather than disease. This study had no pathologic specimens and could not address this question. Dengue was detected in 17/336 (5%) of patients; *Rickettsia* spp. in 6/336 (2%), *Leptospira* spp. in 2/336 (0.6%) and *Coxiella burnetti* and *Brucella* spp. in 1/336 (0.3%) each. The true burden of disease of these pathogens may be higher, given the potential for increased diagnostic yield from serological assays.

The second study[13] is a retrospective analysis of a fever aetiology cohort from Tanzania, in which paired serology for bacterial zoonoses and Rickettsioses was carried out, as well as arboviral PCR. Of 423 enrolled adults, 25 were retrospectively classified as having septic shock, 37 severe respiratory distress without shock and 109 severe pneumonia by WHO Integrated Management of Adolescent and Adult Illness (IMAI) District Clinician Manual criteria[114]. These patients would likely fulfil sepsis criteria under sepsis-2 or 3 guidelines, and were found to have a variety of diagnoses, though not all patients had all diagnostic tests: chikungunya (6/154 [3%]), leptospirosis (5/82 [6%]), Q-fever (7/83 [8%]) and spotted fever group rickettsioses (6/83 [7%]).

### 1.2.6.3 HIV opportunistic infections: PCP, histoplasmosis and cryptococcal disease

The burden of HIV opportunistic infections in sepsis in sSA (including *Pneumocystis jiroveci* pneumonia [PCP], cryptococcal disease and including here histoplasmosis as an opportunistic infection) is unclear. Beyond blood culture identification of *Cryptococcus neoformans* (present in 20/365 of positive blood cultures in the sepsis studies identified in this review) none of these pathogens have been systematically sought in sepsis cohorts in sSA, and their role as causative agents of sepsis is far from clear. Cryptococcal disease most commonly manifests as cryptococcal meningitis, is common in HIV infection and is thought to account globally for 15% of AIDS-related deaths[115]. No study has attempted to define the burden of PCP in sepsis in sSA, though a 2016 systematic review[116] addressed the prevalence and attributable mortality of PCP generally, finding the pooled prevalence of PCP in HIV-infected inpatients (n = 2593, 23 studies) to be 22% (95% CI 17 – 27%) in random effect meta-analysis. Data examining the role of Histoplasmosis as a cause of fever or sepsis in sSA are sparse. A 2015 systematic review[12] identified only one study up to 2013 which Histoplasma urine antigen testing in 628 febrile adults and children in Tanzania finding 9/628 (1%) probable cases, 6/9 of whom were HIV infected. Since then one study in Uganda found 0/151 HIV-infected patients with suspected meningitis[67] had detectable IgM to *Histoplasma capsulatum* and no *Histoplasma* antigen was detected in serum (n = 57), urine (n = 37) or CSF (n=63); a study in Cameroon[68] used histopathologic examination and culture to diagnose histoplasmosis in 7/56 (13%) of HIV infected patients with CD4 < 200 cells $\mu$ L<sup>-1</sup> and chronic cough with histoplasmosis like skin manifestations.

### 1.2.7 Sepsis management

The cornerstone of sepsis management is rapid administration of appropriate antimicrobial therapy, source control of any infectious focus and normalisation of tissue perfusion using intravenous fluids and, if necessary, inotropes, with other organ support as necessary (e.g. intubation and mechanical ventilation and renal replacement therapy). Several international guidelines for sepsis care are available; this section will examine these and specific guidance for sepsis in adults in sSA followed by a review of the evidence to inform these guidelines.

The surviving sepsis campaign has published four editions of comprehensive guidance on the management of sepsis in adults, which are endorsed by all the major critical care organisation in high income settings and form the basis of most sepsis care in high income settings; selected major recommendations of the latest guidance[117] are shown in Table 1.5 below.

Table 1.5: Surviving sepsis campaign guidelines

Recommendation	Strength of recommendation	Quality of evidence
<b>Resuscitation</b>		
Administer 30ml/kg of intravenous crystalloid solution, within 3hr of diagnosis of sepsis	Strong	Low
Use frequent reassessment to guide further fluid	BPS	BPS
Use dynamic variables to assess fluid responsiveness (e.g. cardiac output)	Weak	Low
Use vasopressors in patients who remain hypotensive despite adequate fluid resuscitation; target a MAP of 65mmHg	Strong	Moderate
Use noradrenaline as first-line vasopressor	Strong	Moderate
Measure lactate and use lactate normalisation to guide resuscitation in patients with elevated lactate	Weak	Low
<b>Antimicrobials</b>		
Administer broad spectrum antibiotics within 1hr of diagnosis of sepsis	Strong	Moderate
<b>Adjunctive therapies</b>		
Use hydrocortisone 200mg IV per day if adequate fluid resuscitation and vasopressor therapy are unable to restore haemodynamic stability	Weak	Low
<i>Note:</i> BPS = best practice statement		

Mindful that guidelines aimed at high-income settings may be impossible to implement in low-resource settings (including large areas of sSA) the Global Intensive Care Working Group of the European Society of Intensive Care Medicine (ESICM) published recommendations for sepsis management in resource-limited settings in 2012[118], endorsed by a number of national and international sepsis organisations, and supplements in 2016-17 covering general supportive care[119], infection management[120], management of severe malaria and severe dengue[118] and haemodynamic assessment and support[121] in sepsis in low-resource settings. The major recommendations of this guidance are consolidated in Table 1.6 below.



Table 1.6: ESICM low resource setting sepsis recommendations

Recommendation	Strength of recommendation	Quality of evidence
<b>Resuscitation</b>		
Use capillary refill time, skin mottling scores or skin temperature gradients to assess adequacy of tissue perfusion.	Weak	Ungraded
Use passive leg raise (PLR) to guide fluid resuscitation in sepsis or septic shock	Weak	High
Use crystalloid for fluid resuscitation	Strong	Moderate
Give 30ml/kg of fluid over the first 3hr following sepsis diagnosis, to start within 30mins of recognition	Strong	High
Larger volumes of fluid may be needed if the patient remains fluid responsive and still shows signs of tissue hypoperfusion	Strong	Low
Be extremely cautious in settings with no or limited access to vasopressors and mechanical ventilation and consider stopping fluid if respiratory distress or lung crepitations develop	Strong	High
Use noradrenaline as first line vasopressor	Strong	Moderate
Target a MAP of > 65mmHg	Strong	Moderate
<b>Antimicrobials</b>		
Appropriate antibiotics should be given within the first hour following septic shock	Strong	Low
Source control should occur within 12hr of admission to hospital	Ungraded	Ungraded
<b>Adjunctive therapies</b>		
Use hydrocortisone 200mg IV per day if adequate fluid resuscitation and vasopressor therapy are unable to restore haemodynamic stability	Weak	Low

*Note:*

MAP = Mean arterial blood pressure

The World Health Organisation (WHO) in 2011 published the integrated management of adolescent and adult illness (IMAI) guidance[114], which includes guidance on the management of septic shock and is aimed at district-level clinicians in resource limited settings rather than critical care clinicians. This suggests defining shock as SBP < 90mmHg or pulse > 110/minute and suggest that, once shock is identified, oxygen should be given, a 1 litre bolus of fluid should be given immediately and pulse, SBP and signs of perfusion (urine output, mental status) should be rechecked. If shock persists, another litre should be given; if shock persists after the second litre then help should be sought. Antimicrobials should be administered: the choice of agent should depend on local epidemiology and guidelines, with antimalarials if indicated. No evidence base is referenced for these recommendations.

### 1.2.7.1 Early goal directed therapy

In 2001 a pivotal single centre study in the United States of 263 patients with severe sepsis or septic shock[122] found that protocolised aggressive early resuscitation (called Early Goal Directed Therapy, EGDT) significantly reduced mortality from 46.5% to 30.5% ( $p = 0.009$ ). EGDT called for early central venous catheterisation and protocolised resuscitation to central venous pressure (CVP), MAP and central venous oxygen saturation (ScvO<sub>2</sub>) targets, and was widely adopted. However three large multicentre randomised controlled trials of EGDT – ProCESS in the United States[123], ARISE in Australasia[124] and ProMISe[125] in the United Kingdom, reporting in 2014 and 2015 failed to show any difference in outcomes between the EGDT and usual-care arms. A pre-planned individual level meta-analysis of the 3723 patients included in these trials confirmed similar 90 day mortality in both arms (24.9% for EGDT vs 25.4% for usual care, aOR 0.97 [95% CI 0.82-1.14]) with no benefit found in pre-planned subgroup analysis for patients with worse shock or in hospitals with lower propensity for vasopressors or fluid administration[85]. It is likely therefore that the tenets of EGDT that improve outcomes (early antimicrobials and aggressive fluid resuscitation) have been absorbed into usual care in the fifteen years since the original EGDT study, as evidenced by the reduction in sepsis mortality over this time period, and so the specific package of protocolised care and EGDT targets does not in itself improve outcomes. Unanswered questions now remain regarding the most effective use of the individual components of EGDT (fluids, vasopressors etc.). A number of attempts have been made to develop protocolised sepsis care packages in the style of EGDT for sSA; these are described below in relation to the individual components of sepsis care.

### 1.2.7.2 Evidence to guide antimicrobial therapy in sSA

There is evidence from high income settings that delay in appropriate antimicrobial administration is associated with worse outcomes in sepsis. The first study to investigate this relationship, published in 2006, found a very strong relationship between time to appropriate antimicrobial administration from onset of hypotension and mortality with an absolute increase in mortality of 7.6% for each hour delay over the first six hours[126]. Subsequent data have been more nuanced: a 2015 meta-analysis addressing this question identified 11 studies of 16,178 patients and found no relationship between antimicrobial delay and mortality[127], though many of the included studies are open to confounding by indication (sicker patients are given antimicrobials more quickly), timed antimicrobial administration to non-physiological events (e.g. arrival to hospital or time of blood culture draw rather than onset of hypotension) and did not assess the appropriateness of antimicrobial therapy, all of which could mask a relationship. Appropriate antimicrobial therapy has certainly been shown to be associated with improved survival: a 2010 meta-analysis quantified the

pooled adjusted odds ratio to be 1.6 (95% CI 1.4-1.9) from 26 studies for appropriate versus inappropriate antimicrobial therapy[128]. A recent large retrospective study of 49,331 patients in New York hospitals[7] confirmed the relationship between antimicrobial delay and mortality with an adjusted odds ratio of in-hospital death of 1.04 per hour delay (95% CI, 1.03 – 1.06), and rapid antimicrobial administration forms a key recommendation of current sepsis guidelines.

Data from sSA are lacking, however; neither of the meta analyses above (including between them 37 studies) included any data from sSA, but three of the sepsis studies identified in this systematic review attempt to address the question. The first[17], in an observational study of 382 adults with severe sepsis in Uganda found no association between administration of antibiotics within 1 hour and mortality (OR 0.9 [95% CI 0.6-1.6]) but a total of 42 antibiotic regimens were used and there was a high proportion (22%) of bacteraemic tuberculosis; no assessment of appropriateness of antimicrobials was undertaken and it is possible that inappropriate antimicrobials could mask any association between time of administration and mortality, if one existed.

The second[27], interventional, study in the same centres in Uganda used a before-after design with 661 patients to implement a clinical-officer delivered fluid resuscitation protocol (see below) and administration of antimicrobials. 426 patients were included in the intervention with 245 in the usual care group. The protocol resulted in more rapid administration of antibiotics (67% administered within 1hr versus 30%,  $p < 0.001$ ) and less (though still very prevalent) inappropriate antimicrobial administration (81% versus 95%,  $p < 0.001$ ). Antimicrobial administration was associated with a reduced hazard of death in a multivariable Cox proportional hazards model, but the comparator group used was patients who received no antimicrobials and the hazard ratio for rapid administration ( $< 1$ hr HR 0.44 [95% CI 0.21 – 0.89]) was not significantly different from delayed administration ( $> 6$ hr HR 0.39 [95% CI 0.19 – 0.81]). This type of study design is very prone to bias due to confounding as sepsis management changes over time, especially as the “before” arm was recruited two years before the “after” arm, so results from this study should be interpreted with caution.

A third observational study in a Ugandan teaching hospital[129] provides data on the effect of rapidity of administration of antimicrobials; this study enrolled 218 patients; 89% of them received any antibiotics within 6 hours, with a median time to antibiotic administration of 30mins. Antibiotic administration within 6hr (versus not) was not significantly associated with in hospital mortality in univariate analysis (OR 1.5 95% CI 0.6 – 3.8) though the confidence intervals are wide and could incorporate a clinically significant effect. Again, no assessment of appropriateness of antimicrobials was made.

Only one study provides limited evidence that appropriate antimicrobial therapy improves outcomes in infection in sSA[130]: a combined retrospective-prospective analysis of 104

patients with typhoid perforation (defined by clinical and operative findings rather than culture) from a single Tanzanian teaching hospital found that adequate antimicrobial exposure (defined as at least 3 days of antimicrobial active against *S. Typhi* prior to hospital admission) was associated with improved in-hospital survival in multivariable analysis (aOR 2.9 [95% CI 2.1-4.5]), however it is doubtful that this very specific complication of typhoid fever is generalizable.

### 1.2.7.3 Intravenous fluid therapy in sub-Saharan Africa

The evidence base for rapid fluid administration – and the surviving sepsis recommendation of 30ml/kg within 3hrs following diagnosis - is less secure than for rapid antimicrobial administration. As with antimicrobial administration, adoption of guidelines in response to the EGDT study has meant that disentangling the independent effect of fluid administration is difficult. The data are contradictory. Several large retrospective observational analyses have found no impact on rapidity of fluid bolus administration following sepsis diagnosis[7,131]. Indeed, fluid clearly has the potential for harm; positive fluid balance for patients with sepsis in the ITU has been persistently linked with worse outcomes[7,132,133]. Other studies, in contrast, have found that more rapid initiation of IV fluid is beneficial [134,135]. It may be that heterogeneity in response to fluids plays a role in these conflicting findings; a retrospective multicentre cohort analysis of 3686 patients[136] found that 64% were “fluid responders” – that is, they had a sustained blood pressure response to initial fluid resuscitation without need for vasopressors, and mortality was 15% greater (95% CI 10-18%) in fluid nonresponders.

In sSA, in some ways, the picture is clearer: there is increasing evidence that liberal intravenous fluid administration to septic patients causes harm. The landmark FEAST trial[10] randomised 3141 children with severe febrile illness in Kenya, Uganda and Tanzania to receive either albumin bolus or 0.9% saline bolus or usual care and found an increased risk of death by 48 hours in both bolus groups (RR 1.45 [95% CI 1.13-1.86] for any bolus compared to no bolus). In a secondary analysis[137] this was thought to be due to cardiovascular collapse rather than pulmonary oedema; the mechanism of this is unclear. Only three controlled studies have addressed the question of optimal intravenous fluid resuscitation for septic adults in sSA. One before-after intervention study in septic shock patients carried out in Uganda and described above[27], found a mortality benefit of > 1L fluid over the first 6hr compared to < 1L in multivariable Cox proportional hazard model (HR 0.54 [95% CI 0.35-0.82] 1.0 – 2.5 L vs < 1.0L) but was hampered by a before-after design. Two randomised controlled trials of protocolised early sepsis care in adults have been carried out at a single centre in Zambia assessing the effect of a protocol that administers up to 4L of fluid over 6hrs. The first[25] recruited patients with severe sepsis with organ dysfunction criteria including respiratory rate > 40/min and was stopped early as it was felt

that participants with baseline respiratory compromise might be at risk of harm; 7/10 (80%) of this subgroup died in the usual care group, compared to 8/8 (100%) in the intervention group ( $p = 0.09$ ). The second[9] was unequivocally harmful, with more participants dying by 28 days in the intervention group (48% vs 33%,  $p = 0.03$ ). The reasons are not clear. More respiratory compromise (defined as increase in respiratory rate by 5 breaths/min or reduction in oxygen saturation of 3% or more) occurred in the intervention group (35% vs 22%,  $p = 0.03$ ) but persisted beyond 6 hours in similar numbers in both groups (17% vs 15%,  $p = 0.63$ ).

## 1.3 ESBL-E in sub-Saharan Africa

I here discuss the definition and global epidemiology of ESBL-E, with systematic review and meta analysis of the prevalence of ESBL-E colonisation and invasive infection.

### 1.3.1 Introduction: definition and classification of ESBL-E

$\beta$ -lactamases are enzymes that hydrolyse the active  $\beta$ -lactam ring in  $\beta$ -lactam antimicrobials. Two classification schemes are usually used for  $\beta$ -lactamases: the molecular (or structural) classification of Ambler[138], or the Bush-Jacoby-Medeiros functional classification[139] (Table 1.7). Molecular classification is straightforward and depends on protein homology; class A, C and D enzymes are serine  $\beta$ -lactamases and class B are metallo- $\beta$ -lactamases, named for the composition of their active site. The functional classification is complex and clusters enzymes into four groups, with a number of subgroups, based on substrates and the effect of  $\beta$ -lactamase inhibitors and EDTA: class 1 (corresponding to Ambler class C) are cephalosporinases that are not inhibited by clavulanic acid, and includes the AmpC enzymes of the Enterobacteriaceae; class 2 enzymes are  $\beta$ -lactamases that are largely inhibited by clavulanic acid and belong to Ambler class A or C; and class 3 are the metallo- $\beta$ -lactamases corresponding to Ambler class B. Class 4 enzymes are penicillinases which are not inhibited by clavulanic acid, though are of limited significance and not included in Table 1.7.

The vast majority of clinically relevant ESBLs belong to Ambler class A, functional class 2be. For the purpose of this thesis, therefore, I define ESBL as enzymes which confer resistance via hydrolysis to penicillins, cephalosporins of the first, second or third generation (excluding cephamycins), aztreonam, but not carbapenems, and are inhibited by  $\beta$ -lactamase inhibitors such as clavulanic acid[140]; this corresponds to the Bush-Jacoby group 2be and makes clear that I draw a distinction between ESBL and AmpC enzymes, which would be grouped with the Bush-Jacoby group 1. When referring to ESBL genes I will refer to them in the standard way using *bla* and a subscript to indicate the gene as for example *bla*<sub>CTX-M-15</sub>; when referring to the enzyme, I will refer to them as, for example CTX-M-15.

Table 1.7: ESBL classification. Adapted from [139]

Bush Jacoby group	Molecular class	Distinctive substrates	Inhibited by		Defining hydrolysis spectrum characteristics	Representative enzymes
			BLI	EDTA		
1	C	Cephalosporins	No	No	cephalosporins > benpen, hydrolyzes cephamycins	E. coli AmpC, P99, ACT-1, CMY-2, FOX-1, MIR-1
1e	C	Cephalosporins	No	No	ceftazidime and often other oxymino-beta-lactams	GC1, CMY-37
2a	A	Penicillins	Yes	No	benzylpen > cephalosporins	PC1
2b	A	Penicillins, early cephalosporins	Yes	No	Similar hydrolysis of benzylpenicillin, cephalosporins	TEM-1, TEM-2, SHV-1
2be	A	Extended- spectrum cephalosporins, monobac- tams	Yes	No	oxymino-beta lactams	TEM-3, SHV-2, CTX-M-15, PER-1, VEB-1
2br	A	Penicillins	No	No	Resistance to BLI	TEM-30, SHV-10
2ber	A	Extended- spectrum cephalosporins, monobac- tams	No	No	oxymino-beta lactams plus resistance to BLI	TEM-50
2c	A	Carbenicillin	Yes	No	Increased hydrolysis of carbenicillin	PSE-1, CARB-3
2ce	A	Carbenicillin, cefepime	Yes	No	Increased hydrolysis of carbenicillin, cefepime, and ceftiofame	RTG-4
2d	D	Cloxacillin	Variable	No	Increased hydrolysis of cloxacillin or oxacillin	OXA-1, OXA-10
2de	D	Extended- spectrum cephalosporins	Variable	No	cloxacillin or oxacillin and oxymino-beta-lactams	OXA-11, OXA-15
2df	D	Carbapenems	Variable	No	cloxacillin or oxacillin and carbapenems	OXA-23, OXA-48
2e	A	Extended- spectrum cephalosporins	Yes	No	Inhibited by clavulanic acid but not aztreonam	CepA

Table 1.7: ESBL classification. Adapted from [139] (*continued*)

Bush Jacoby group	Molecular class	Distinctive substrates	BLI	EDTA	Defining hydrolysis spectrum characteristics	Representative enzymes
2f	A	Carbapenems	Variable	No	carbapenems, oxymino-beta- lactams, cephamycins	KPC-2, IMI-1, SME-1
3a	B (B1)	Carbapenems	No	Yes	includes carbapenems but not monobactams	IMP-1, VIM-1, CcrA, IND-1
	B (B3)					L1, CAU-1, GOB-1, FEZ-1
3b	B (B2)	Carbapenems	No	Yes	Preferential hydrolysis of carbapenems	CphA, Sfh-1

*Note:*

BLI = Beta-lactamase inhibitor

### 1.3.2 Global molecular epidemiology of ESBL-E: an overview

The history of the global spread of ESBL-E is complex and an enormous number of unique ESBL amino acid sequences have been described; at the time of writing the NCBI beta-lactamase directory contains 1557 named  $\beta$ -lactamase genes, many of them ESBL. However, there are 3 families of enzymes which cause the majority of infections in humans: TEM, SHV, and CTX-M. They will be briefly described here in turn in the context of their putative origins and global dissemination in the latter half of the 20th century. A diverse range of other ESBL enzymes have been described, but are largely of less clinical significance than those described above, and are beyond the scope of this review: most notably the OXA type, which in contrast to TEM, SHV, and CTX-M, are of the molecular class D and functional class 2d, and are characterised by a high rate of hydrolysis of cloxacillin[141]; like TEM and SHV, OXA  $\beta$ -lactamases are not always extended-spectrum.

#### 1.3.2.1 1980s-1990s: First identification of ESBL in nosocomial pathogens

$\beta$ -lactamases form an integral part of the natural armamentarium of many genera of bacteria – particularly Gram negatives, including Enterobacteriaceae - and predate the antibiotic era; penicillinases were identified in *E. coli*, for example, prior to the widespread introduction of penicillin for treatment of human disease[142]. These  $\beta$ -lactamases are often chromosomally located; the first plasmid-mediated narrow-spectrum  $\beta$ -lactamase, TEM-1 -named for the patient, Temoneira, from whose blood it was first isolated – was found in Athens in the 1960s[143]. It rapidly disseminated globally and is thought to be responsible

for a high proportion of ampicillin resistance in *E. coli*[141]. This worldwide spread spurred the development and use of  $\beta$ -lactamase resistant extended-spectrum cephalosporin antimicrobials, which found wide use in the 1980s. Perhaps inevitably, an enzyme conferring resistance to extended-spectrum oxyimino-cephalosporins was subsequently identified in a German clinical *Klebsiella ozaenae* isolate in 1983, encoded for on a pBP60 plasmid and enzymes of this sort were named ESBLs[144,145].

This first ESBL enzyme was found to be similar to an existing plasmid-encoded narrow spectrum beta lactamase, SHV-1, which had been described in the 1970s in *E. coli*, and was thought to itself be descended from a chromosomally encoded *K. pneumoniae* narrow spectrum beta lactamase which was liberated onto a plasmid[146]. The point mutations in SHV-1 conferred the ESBL phenotype, and this enzyme was named SHV-2. This pattern - mutation of a narrow spectrum  $\beta$ -lactamase to produce an ESBL phenotype - also occurred in TEM, and the first ESBL TEM was described in France in 1989[147] and named TEM-3. Many TEM and SHV variants were subsequently described[148]. However, in this early stage of the epidemic, ESBL enzymes were largely nosocomial, and often associated with *Klebsiella* spp.[149].

### 1.3.2.2 1990s-2010s: Emergence and globalisation of CTX-M

From the late 1990s onwards, there were profound changes in the global epidemiology of ESBL-E, on three fronts, all intricately interrelated, and occurring simultaneously: first, the rapid emergence and globalisation of the successful CTX-M ESBL enzyme family[150], aided by mobile genetic elements; second, *E. coli* joining *Klebsiella* spp. as a major ESBL host[151], and the emergence of so-called high risk bacterial clones; and third, the spread of ESBL-E into the community[152]. CTX-M-1 was first identified and named in Germany in 1989[153], the name derived from “active on cefotaxime first isolated in Munich.” Many variants were subsequently identified, largely in *E. coli* and *K. pneumoniae*, from isolates all over the world[154]. The *bla*<sub>CTX-M</sub> genes are clustered by homology into 5 groups (*bla*<sub>CTX-M</sub> groups 1,2,8,9 and 25) and each group is thought to have descended from a chromosomal beta lactamase from *Kluyvera* spp.[150]

A year-on-year rise in incidence of invasive ESBL-E infection was seen in most high-income settings (Figure 1.2) throughout the 2000s and 2010s, the majority of which were CTX-M producers, though with varying proportions of different CTX-M enzymes in different locations[151,155]. Risk factors for ESBL-E infection in high income settings have persistently been shown to be hospital or long-term care facility exposure, antimicrobial exposure and chronic health conditions though it was recognised in the 2000s that a large proportion of patients with invasive ESBL-E lack any of these risk factors[152], suggesting acquisition in the community. Colonisation prior to infection is thought to be the norm;



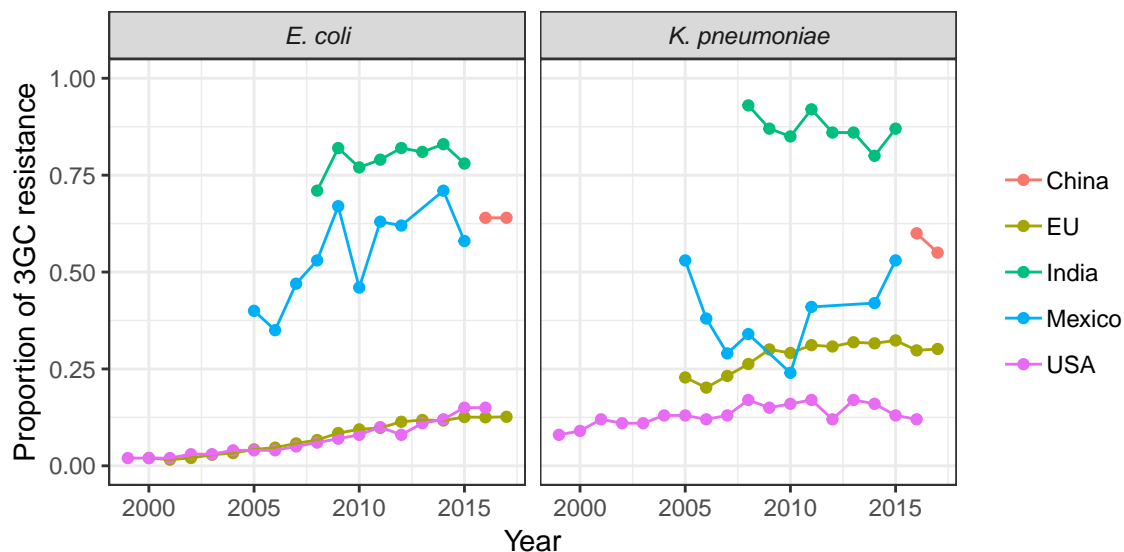


Figure 1.2: Prevalence of third generation cephalosporin resistance in representative high income (EU, USA) and middle income (China, Mexico, India) areas in invasive *E. coli* (left panel) and *K. pneumoniae* (right panel) isolates. Data for EU are from ECDC surveillance atlas (<https://ecdc.europa.eu/en/antimicrobial-resistance/surveillance-and-disease-data/data-ecdc>) and for other countries are from CCDEP resistance map (<https://resistancemap.cddep.org/AntibioticResistance.php>) both accessed 28 December 2018. 3GC = third generation cephalosporin. Note that these data are 3GC-resistant isolates rather than confirmed ESBL-producers, but would be expected to be ESBL-producers in the vast majority of cases.

prior colonisation is a significant risk factor for infection and indeed when sought ESBL-E are found in the stool of healthy community members worldwide[156].

Though less comprehensive, data from middle income countries suggests that prevalence of ESBL producers amongst invasive *E. coli* and *K. pneumoniae* can be very high (Figure 1.2) and in countries such as India invasive *E. coli* and *K. pneumoniae* that are sensitive to third-generation cephalosporins are in the minority. The reasons for this are not clear but country and regional level associations (which are open to ecologic bias) have been shown with antimicrobial consumption[157,158] and economic status; GDP per capita has been found to correlate inversely at a country level with third-generation cephalosporin resistance rates[159]. Data from sSA have historically been lacking and are systematically reviewed below.

### 1.3.2.3 Epidemiology of gut mucosal carriage of ESBL-E: the first step towards invasive infection

Invasive infections with Enterobacteriaceae are thought to usually result from infection from an individual's own gut microbiota, irrespective of resistance pattern[160], and whole genome sequencing has confirmed that invasive isolates are often closely related to prior gut

carriage isolates[161]. Strategies to minimise carriage are therefore potentially attractive as interventions to reduce invasive infection and a number of studies have attempted to understand the dynamics of gut mucosal ESBL-E carriage in health and disease. A brief overview of ESBL-E community carriage is presented here, and ESBL-E carriage in sSA specifically is systematically reviewed below.

ESBL-E community carriage in Spanish outpatients[162] and healthy children in Poland[163] was first described in 2001. Since that time carriage by healthy community members has been found worldwide in all populations[156,164], though there are heterogeneities between and within countries which mirror the prevalence of invasive ESBL-E prevalence amongst *E. coli* and *K. pneumoniae*. In Europe, for example, community prevalence of ESBL-E carriage was estimated be 7.3% in the UK in 2014 in a large community study[165], 4.5% in the Netherlands in 2012[166] and 4.7% in Sweden in 2012/13[167] and 3.7% in Spain in 2003[168], significantly lower than community carriage prevalence of 50.9% seen in China in 2009[169] or 33.8% in India in 2011-2013[170].

Risk factors for colonisation have been identified in many studies and antimicrobial exposure[171,172] and healthcare facility exposure[170,173] (including long term care facilities[174]) are consistently identified as such. Colonisation of a household member has also been identified as a risk factor[175,176], suggesting significant within-household spread. Antacid use has been associated with ESBL-E colonisation[171] as has exposure to farming[166]. In low prevalence areas, travel to high prevalence areas is a risk factor[165,167,171,172,177].

The majority of studies of ESBL-E carriage are cross sectional and only a handful have attempted to characterise longitudinal carriage of ESBL-E with a longitudinal sampling approach. Estimates of carriage duration vary, partly because of the difficulty in inferring them from interval-censored rectal swab or stool data, but it is clear that some patients remain colonised for many months. Following a Swedish ESBL-E outbreak, 12% of patients still carried ESBL-E at the final sampling visit, a median 58 months after the outbreak[178]. French and German studies found a median duration of carriage of 4.3[179] and 12.5[180] months respectively following hospitalisation or outbreak. More transient carriage following international travel seems to be the norm with a median of 30 days in a large Dutch study[181]; the reasons for this are not clear.

The largest longitudinal community study of ESBL-E carriage took place in the Netherlands which recruited 76 ESBL-E colonised and 249 uncolonised community members and carried out longitudinal stool sampling at 5 time points over 8 months. 25/76 (33%) of initially-colonised participants remained persistently colonised after a median 242 days. Antimicrobial exposure in the past 6 months, proton-pump inhibitor use, colonisation with *E. coli* phylogroup B2 or D and presence of *bla*<sub>CTX-M-27</sub> or *bla*<sub>CTX-M-14</sub> was associated with persistent carriage, suggesting both host and bacterial factors may be important

determinants of carriage duration. *K. pneumoniae* colonisation seemed to be less common in the persistent carriage group[177,182]. This study also found significant heterogeneity of *E.coli* sequence type in longitudinal samples of persistent carriers but that ESBL genes and often detectable plasmid replicons remained unchanged, suggesting a significant role for mobile genetic elements.

#### 1.3.2.4 Molecular mechanisms underlying success of CTX-M: mobile genetic elements and high-risk clones

The remarkable success of CTX-M has led to efforts to understand the molecular mechanisms by which this enzyme spread so rapidly. The system is complex, and poorly understood, but should be considered at multiple levels including that of the organism; the plasmid; the transposon, which may contain integrons or insertion sequences and, at the lowest level the ESBL gene. These will briefly be reviewed here.

The initial mobilisation event of *bla*<sub>CTX-M</sub> from *Kluyvera* spp. is thought to have been mediated by capture of transposable insertion sequences; the insertion sequence *ISEcp1* has been experimentally demonstrated to mobilise the *bla*<sub>CTX-M</sub> precursor from *Kluyvera ascorbata*[183] and *ISEcp1* is most consistently associated with *bla*<sub>CTX-M</sub> genes. However, *IS26*, *ISCR1* and *IS10* have also persistently been described upstream from *bla*<sub>CTX-M</sub> genes, suggesting multiple mobilisation events[155]. *bla*<sub>CTX-M</sub> genes cluster by homology into groups and there is an association between particular *bla*<sub>CTX-M</sub> gene clusters and particular insertion sequences, consistent with a hypothesis of multiple mobilisation events[184]. These insertion sequences provide two roles: they encode a transposase enabling gene mobilisation but act as a strong promotor of *bla*<sub>CTX-M</sub>, without which phenotypic cephalosporin resistance is absent or reduced[185].

After mobilisation from the *Kluyvera* genome, the *bla*<sub>CTX-M</sub> genes were integrated onto a plasmid backbone, a process which is likely ongoing as a substantial number of diverse *bla*<sub>CTX-M</sub> carrying plasmids have been described: there is, however, an association between *bla*<sub>CTX-M</sub> genotype and plasmid incompatibility group. The successful *bla*<sub>CTX-M-15</sub> gene is very strongly associated with the narrow host-range IncF plasmid group, for example, which are restricted to Enterobacteriaceae[185,186]. Identical *bla*<sub>CTX-M</sub> containing plasmids have been found across diverse geographical regions and have been termed “epidemic plasmids”[155] though the mechanism of persistence of these plasmids within a bacterial population remains unclear.

In addition to frequently co-occurring *bla*<sub>CTX-M</sub> genes, transposable elements and plasmids, some clonal groups of *E. coli* and *K. pneumoniae* are both globally successful and associated with particular *bla*<sub>CTX-M</sub> genes and plasmids. These successful sequence types (STs) are known as “high risk clones.” The archetypal example is *E. coli* ST131 which is often

associated with an IncFII plasmid containing *bla<sub>CTX-M-15</sub>*[187]. First described in 2008, *E. coli* ST131 is thought to be responsible for around 80% of extra-intestinal ESBL *E. coli* infection[188]. Population genomics studies have demonstrated that a particular clade, ST131 clade C, is globally dominant and have shown a sequential acquisition of virulence determinants followed by mobile genetic elements conferring fluoroquinolone and ESBL resistance[189,190]. These events may have contributed to the global success of ST131, but the precise mechanism of its apparent fitness advantage remains unknown.

### 1.3.3 Epidemiology of ESBL-E in sub-Saharan Africa

In order to clearly define the epidemiology of ESBL-E in sSA I performed a systematic review and meta analysis, which is presented here.

#### 1.3.3.1 Search strategy

A systematic review of the literature was undertaken to answer the following questions: firstly, what is the prevalence of ESBL-E amongst invasive isolates of *Klebsiella pneumoniae* and *Escherichia coli* infecting humans in sub-Saharan Africa? Secondly, what is the prevalence of gut mucosal carriage of ESBL-E amongst humans in sSA, and what risk factors for carriage have been identified? To this end a search of PubMed and Scopus was carried out using the search terms given in the chapter appendix.

Inclusion criteria were any study that took place in sSA and allowed the calculation of a prevalence of ESBL-producers in *K. pneumoniae* or *E. coli* amongst invasive human isolates, or prevalence of human gut mucosal carriage of ESBL-E. Studies were excluded if no ESBL-E confirmatory testing was performed using phenotypic (double disc or combination disc or E-test) or molecular (PCR or whole genome sequencing) methods. Invasive isolates were defined to be any blood or CSF sample or other usually sterile fluid, or urine or wound swabs with clinical suspicion of infection.

#### 1.3.3.2 Statistical analysis

Data were extracted from the identified studies: prevalence of ESBL-E in *K. pneumoniae* or *E. coli* amongst invasive human isolates, or prevalence of human gut mucosal carriage of ESBL-E. Proportions were plotted in forest plots with exact binomial confidence intervals, and stratified by location of isolation in the case of carriage isolates (community, outpatient, on hospital admission, or inpatient). Summary estimates were calculated using random-effect meta analysis with generalised linear mixed models (GLMMs) assuming binomially distributed prevalences and normally distributed random effects - the normal-binomial

model - using the R packages *meta*[69] v4.9.5 and *lme4* v1.1.21[191]. Heterogeneity was explored by meta-regression, particularly regressing proportion of ESBL producing *E. coli* and *K. pneumoniae* against year of isolation. This was achieved by adding year as a fixed effect covariate and assessing explanatory power of a model including the parameter to one without using likelihood ratio testing and considering  $p < 0.05$  to be statistically significantly better fit. Predicted population prevalences were generated from the fitted models and plotted with 95% confidence intervals generated from 1000 bootstrap re-samples. Where available, data on the identified ESBL enzymes were also extracted and plotted as simple proportions.

### 1.3.3.3 Results

On 8th December 2018 the search identified 2975 unique studies; after abstract review 192 underwent full-text review, resulting in the inclusion of 86 studies, 54[192–244] providing data on invasive infection and 32 [245–277] on carriage. Details of these studies are given below and in the chapter appendix.

### 1.3.3.4 Invasive ESBL-E infection

Table 1.11 in the chapter appendix shows the 54 included studies in this analysis, which provide data on 6067 *E. coli* and 2974 *K. pneumoniae* isolates. All studies were cross sectional in design. Of the 54, 18/54 were laboratory based (i.e. a survey of all samples received in the laboratory); 17/54 were truly invasive in that they included predominantly blood culture; a combination of urine, CSF, and wound swabs were included in the remaining studies. 36/54 studies provided data on adults and children; 6/54 on adults only; and 12/54 on children only. The majority of studies (42/54) include both community and nosocomial acquired infection. Of the remainder, 3/54 provided data on nosocomial infection only. Figure 1.3A shows a map of available data by country; data are available from across the continent though Nigeria (8 studies) and Tanzania (7 studies) are over represented and many countries provide no data.

The proportion of ESBL producers amongst invasive *E. coli* and *K. pneumoniae* in sSA is heterogeneous but many studies show an extremely high prevalence (Figures 1.3B and 1.3C), comparable to that seen in the Indian subcontinent and other high-prevalence areas and highlighting the scale of the public health problem posed by ESBL-E in sSA. Meta regression shows clear temporal trends of an increase over time: addition of time as a fixed-effect covariate to the random effects model gives improved fit on likelihood ratio testing of nested models ( $p < 0.001$  for both *E. coli* and *K. pneumoniae*). Though data are sparse pre-2000, those data that are available suggest that ESBL producing *E. coli* and *K. pneumoniae* were identified in West Africa even in the 1990s: a retrospective laboratory

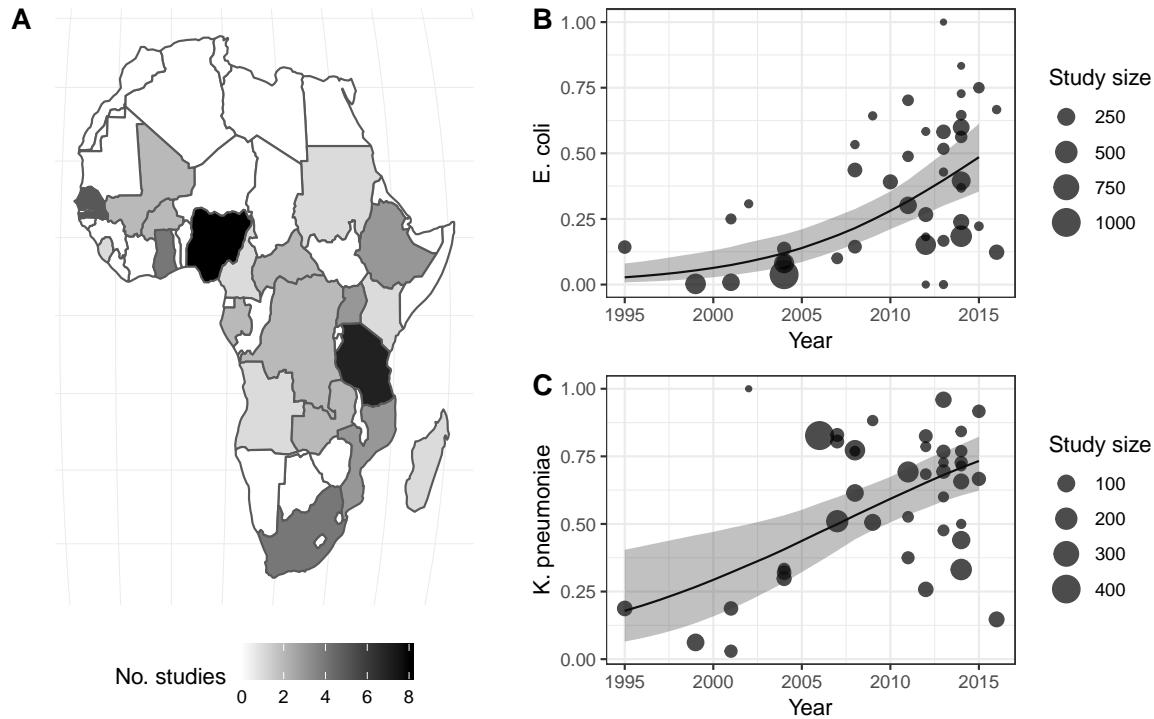


Figure 1.3: Invasive ESBL-E in sSA. A: Available studies by country. B and C: meta regression of proportion of invasive *E. coli* and *K. pneumoniae* respectively as a function of time. In both cases time is statistically significantly associated with proportion of ESBL ( $p < 0.001$  on likelihood ratio testing of nested models). 95% CI generated from 1000 bootstrap replicates.

based study in Yaounde, Cameroon on isolates from a variety of clinical samples from 1995-1998 found that 13/91 *E. coli* and 12/64 *K. pneumoniae* were ESBL producers, with the *bla<sub>SHV-12</sub>* gene predominant[236]; in Dakar, Senegal, 6/97 *K. pneumoniae* isolates from community acquired urinary tract infections in 1999-2000 were found to be ESBL producers[239].

Some of the heterogeneity in prevalence does however seem to be explained by sample type; a clearer picture appears when the analysis is restricted to the 16 studies including predominantly blood culture (Figure 1.4)[196,198–203,205,208,209,211,213,218,221,235,244]. In this analysis it seems clear that the worldwide epidemiology of ESBL-E was mirrored in sSA; ESBL initially spread amongst invasive *K. pneumoniae* post 2000 (particularly post 2005) before becoming established in *E. coli* after 2010. In 2014, the latest available data, the pooled population prevalence of ESBL from binomial-normal random effects meta analysis was 61% [95% CI 40-80%] amongst *E. coli* and 86% [95% CI 73-92%] amongst *K. pneumoniae* bloodstream infection isolates, suggesting endemicity of ESBL amongst these pathogens in sSA, and comparable to the highest prevalence areas in the world.

ESBL genes were characterised in 10 studies by whole genome sequencing[244] ( $n=1$ ) or by PCR[197,200,201,203,210,218,219,235,238] ( $n=9$ ) for 821 *E. coli* and 791 *K. pneumoniae* isolates (Figure 1.5). *bla<sub>CTX-M</sub>* were the most commonly occurring ESBL genes, and

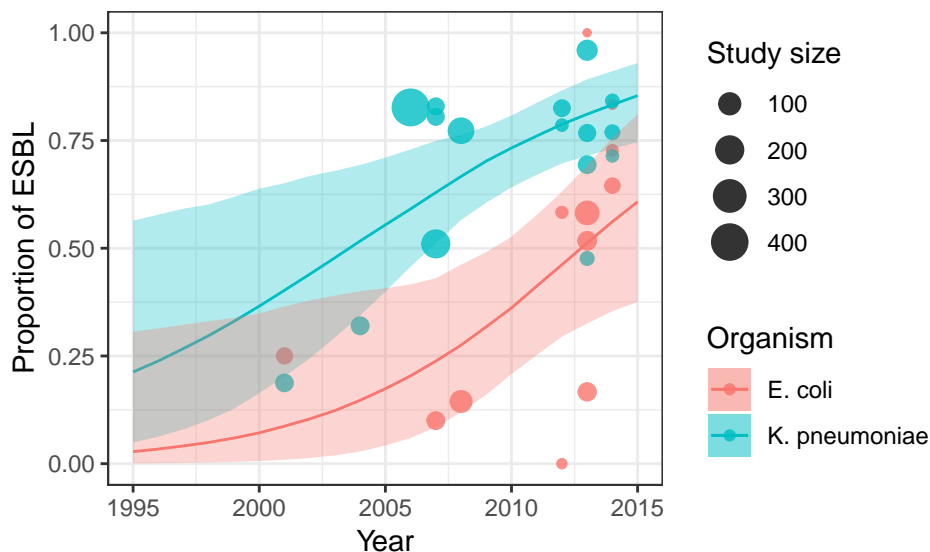


Figure 1.4: Meta regression of proportion of ESBL producing *E. coli* and *K. pneumoniae* amongst invasive isolate in sSA from studies carrying out blood culture, as a function of time. Includes 1242 *K. pneumoniae* and 489 *E. coli* isolates. 95% CI generated from 1000 bootstrap replicates from fitted models. In both cases time is statistically significantly associated with proportion of ESBL ( $p < 0.001$  on likelihood ratio testing of nested models).

the majority of these were *bla*<sub>CTX-M-15</sub> in both organisms. *bla*<sub>OXA</sub>, *bla*<sub>TEM</sub> and *bla*<sub>SHV</sub> genes were also commonly found but were often not further characterised, presenting some problems of interpretation, as these enzymes can be narrow or broad-spectrum beta-lactamases. Certainly, *bla*<sub>SHV-1</sub> and *bla*<sub>TEM-1</sub> encode narrow spectrum beta lactamase enzymes, which were commonly identified in these studies, though only a handful of isolates had characterisation of *bla*<sub>SHV</sub> genes beyond identification of the *bla*<sub>SHV</sub> group. All the identified *bla*<sub>OXA</sub> genes were narrow spectrum beta lactamases (*bla*<sub>OXA-1</sub> or *bla*<sub>OXA-30</sub>). These data suggest that the genomic landscape of invasive ESBL-E in sSA is dominated by *bla*<sub>CTX-M</sub>, and *bla*<sub>CTX-M-15</sub> in particular, mirroring that seen worldwide.

Though no data were identified from Malawi that fulfilled the inclusion criteria of the systematic review, there are three studies that suggest the epidemiology of invasive ESBL-E in Malawi is similar to that described above. A study from Blantyre in 2004-2005 found that ESBL-E were unusual in blood stream infection (BSI) isolates: of 1191 Enterobacteriaceae BSI, only 8 unique isolates showed an ESBL phenotype (*K. pneumoniae* 4/8, *K. oxytoca*, 1/8, *Enterobacter cloacae* 2/8 and *E. coli* 1/8) though no denominators are provided to allow calculation of prevalence. In terms of ESBL enzymes, CTX-M-15 ( $n = 1$ ) was described, though in the minority: SHV-11 ( $n = 1$ ), SHV-12 ( $n = 3$ ), SHV-27 ( $n = 1$ ) and TEM-63 ( $n = 2$ ) were the other enzymes identified[278]. Longitudinal blood culture surveillance in Blantyre suggests that after 2005 – which coincided with the introduction of ceftriaxone in government hospitals – the prevalence of ceftriaxone resistance rapidly increased, to 90.5% in *K. pneumoniae* and 30.3% in *E. coli* BSI isolates by 2016[15], though this study

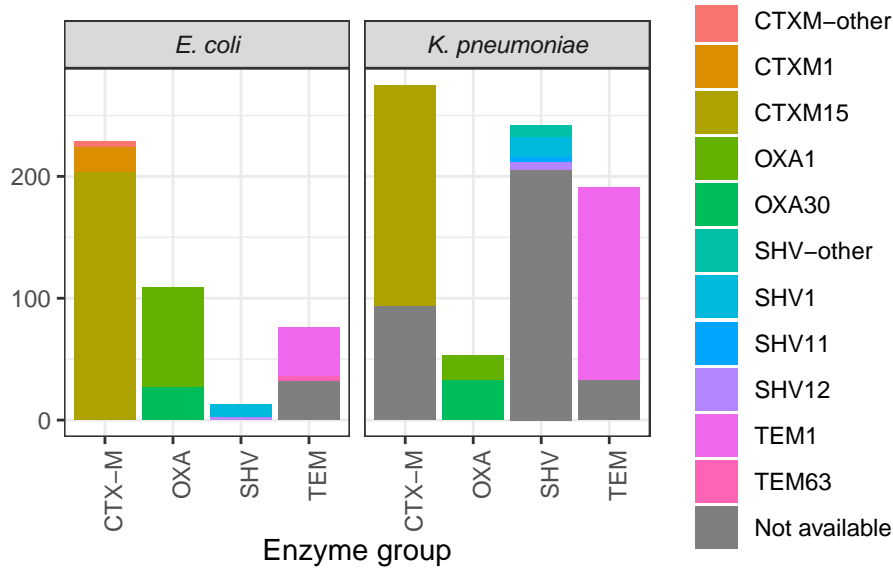


Figure 1.5: Distribution of beta-lactamase enzymes in invasive ESBL producing *E. coli* (n= 821) and *K. pneumoniae* (n = 791) from 10 studies. Invasive is here defined as those recovered from sterile sites or from urine or wound swabs with clinical suspicion of infection.

did not carry out confirmatory ESBL testing. Finally, two retrospective whole-genome sequencing study which selected 94 diverse (largely invasive) clinical *E. coli* and 72 *K. pneumoniae* isolates from Blantyre from 1996-2014 found that 21/94 *E. coli* isolates carried an ESBL gene, with CTX-M predominating (20/21)[279] and 31/60 *K. pneumoniae* had an ESBL phenotype again with *bla*<sub>CTX-M-15</sub> predominant (39% [28/72] of identified ESBL genes)[280].

### 1.3.3.5 Gut mucosal carriage of ESBL-E in sub-Saharan Africa

Table 1.12 in the chapter appendix shows the 32 identified studies that provide data on gut mucosal carriage in different populations in sSA. The populations recruited to the studies are heterogeneous but include community members, hospitalised patients, outpatients, orphanage residents, hospital workers and food handlers in schools. Adults and children are included. Data on 10,232 individuals from 19 countries are available in total, and are plotted in Figure 1.6. The earliest samples were collected from staff and children in a Malian orphanage in 2003, when 72% (49/68) of participants were found to be colonised with ESBL-E[271]. There is significant heterogeneity in prevalence, some of which is explained by the study population (Figure 1.6); inpatients tend to have a higher ESBL-E carriage prevalence than community members. Outpatients have similar carriage prevalence to community members but inpatients even on hospital admission seem to have a higher carriage prevalence than community members.



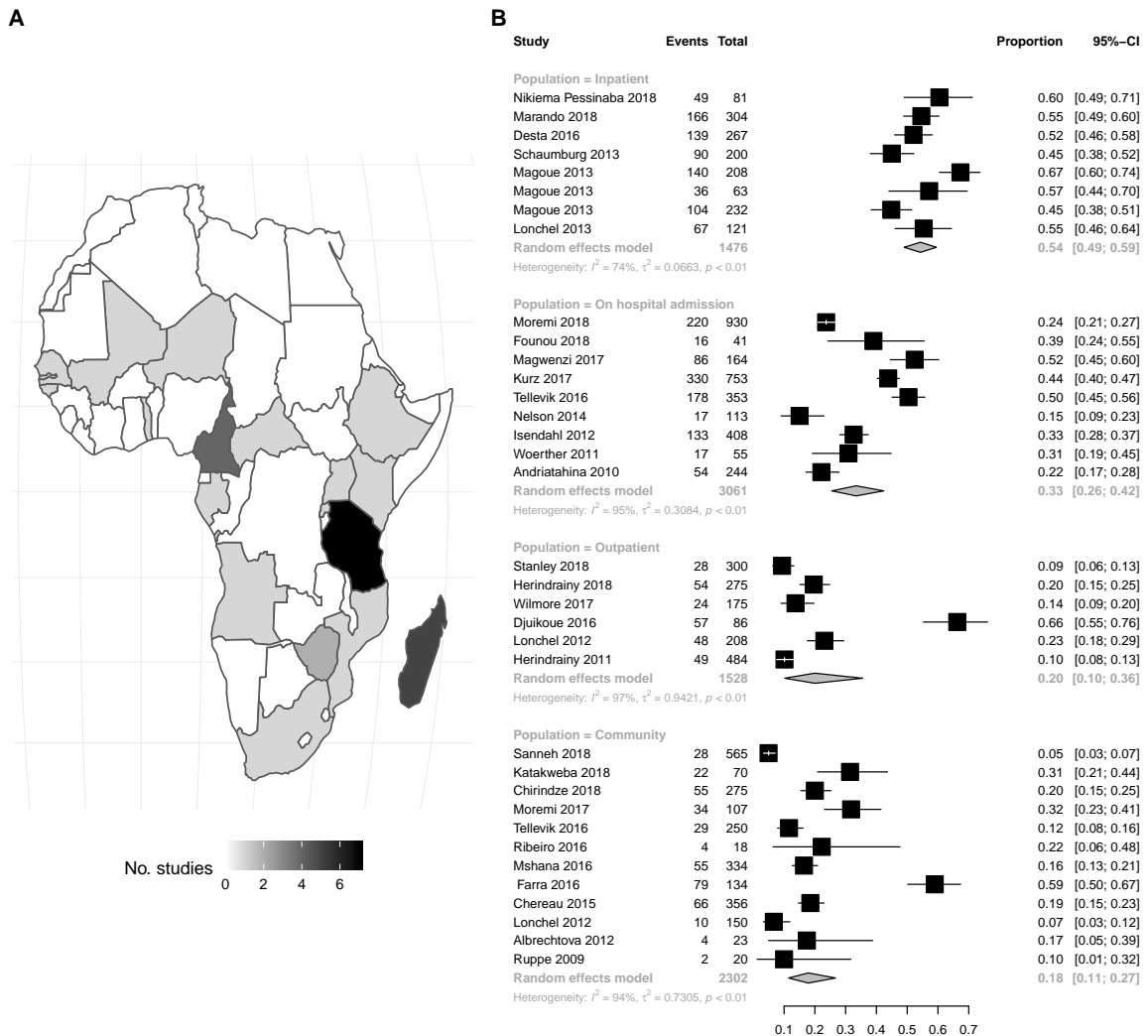


Figure 1.6: ESBL-E gut mucosal carriage in sSA. A: included studies by country; B: forest plot of ESBL-E carriage prevalence stratified by population. Pooled random effect summary estimates shown.

Significant heterogeneity in prevalence persists across all groups meaning that summary estimates should be interpreted with caution; community carriage of ESBL-E ranges from 5% in adults in The Gambia in 2015[275] to 59% in children in the Central African Republic in 2013[250], but a summary estimate from a random effect meta analysis is that 18% (95% CI 12-28%) of community members in sSA are colonised with ESBL-E, significantly higher than the prevalence in high-income settings.

Hospitalisation is clearly a driver of ESBL-E colonisation in the included studies - hospitalised cohorts have persistently higher prevalence of ESBL-E carriage – and prior antimicrobial exposure is consistently identified as a risk factor for carriage[252,259,275]. Consistent with a putative faecal-oral transmission route, boiling water and using a borehole as a source of water were identified as protective factors in studies in Rwanda[274] and Togo[277]

respectively. Data to elucidate the role of within-household transmission are sparse, though one study in Rwanda found that a colonised family member was independently associated with ESBL-E carriage on admission to hospital[277]. Lower socioeconomic status was found to both be protective against ESBL-E colonisation in the Central African Republic[250] and be associated with ESBL-E colonisation in Madagascar[266]; this relationship is likely to be complex and mediated by, for example, local availability and cost of antimicrobials. The role of HIV is not clear: in children in Dar-es-salaam, Tanzania, ESBL-E carriage was much more common amongst HIV infected children[255], and in Harare Zimbabwe, receipt of ART for less than a year was associated with carriage[249]. This relationship is very open to confounding and many studies have not found an association between ESBL-E carriage and HIV infection[252,259,265,273,275,277].

Data on ESBL enzymes present in carriage isolates are available for 996 *E. coli* and 607 *K. pneumoniae* from 8 studies (Figure 1.7), showing a similar picture to invasive isolates; the landscape is dominated by CTX-M-15. One study used whole-genome sequencing [272], the remainder used a variety of PCR techniques[248,260–262,265,267,273].

Only 4 studies are longitudinal cohorts which could provide insight into temporal trends and determinants of carriage[267,269,273,277]; all of these studies were health facility based and ascertained ESBL status on admission and discharge. Significant increases in ESBL-E carriage were seen in all studies: from 50% to 65% in Rwanda; from 30% to 95% in Niger; from 21% to 57% in Madagascar; and from 23% to 36% in Tanzania. No studies followed patients into the community, thus carriage duration of ESBL-E in sSA remains unknown and no interventional studies identified aiming to reduce ESBL-E carriage were identified.

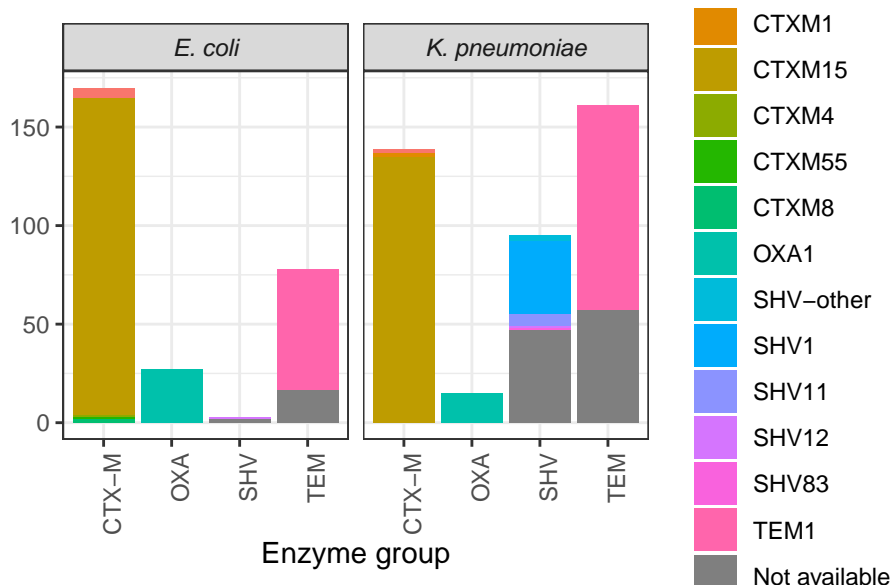


Figure 1.7: Distribution of beta-lactamase genes in carriage ESBL producing *E. coli* (n = 996) and *K. pneumoniae* (n = 607) from 9 studies. Carriage isolates are here defined as those isolated from stool or rectal swabs.

## 1.4 Conclusions

The aetiology of sepsis in sSA is poorly defined, hence optimal antimicrobial strategies are unknown; disseminated TB is likely to play a significant role, but data to guide tuberculosis therapy strategies in the critically unwell are lacking. The role of bacterial zoonoses, arboviruses and HIV opportunistic infections are not well defined, but may be significant. Diagnostic uncertainty and paucity of microbiologic support across sSA may be creating a permissive environment for the widespread broad spectrum antimicrobial use, often third-generation cephalosporins. It is likely that dose, duration and indication are frequently inappropriate, and thus could contribute both to increased mortality and to spread of ESBL-E.

Certainly, ESBL-E are endemic in sSA and are a problem of serious public health concern; sSA has rates of ESBL-E in invasive disease that are comparable to the highest in the world and ESBL-E gut mucosal carriage in healthy populations across the continent is common. Whilst it is clear that health care facilities are strongly associated with ESBL-E acquisition, a deeper understanding of the determinants and sources of acquisition, and carriage duration is lacking. In order to understand the role of health facilities in driving the ESBL-E pandemic, a high quality longitudinal ESBL-E carriage data from both healthy and sick (health facility exposed) populations are required.

I propose that optimising the management of severe febrile illness can tackle two problems: reduce over prescription of broad spectrum antimicrobials and improve outcomes in critically unwell patients. This may seem counter-intuitive: why target the very unwell with an antimicrobial stewardship intervention? However I hypothesise that the “step-up” way in which we approach antimicrobial therapy in the immunosuppressed in resource limited settings is flawed, is driving the twin problems of poor sepsis outcomes and AMR, and may represent a low hanging fruit to tackle both problems. In a setting where so much management is empiric, current management begins with broad spectrum antibacterials and adds in further therapies - TB therapy, PCP therapy, or antifungals - based on non-response. This is the management that is codified in the WHO guidance for treating critically unwell TB suspects, but I suggest that it results in prolonged antibacterial exposure and delay in definitive treatment. We can imagine an alternate therapy, whereby we start broad, and rapidly narrow the spectrum of therapy based on the results of investigations. But such a strategy requires data; what are the causes of sepsis in sSA that we should target? How should we rationalise therapy? What are the determinants of AMR acquisition in sepsis survivors and how can we mitigate against acquisition? It is the aim of this thesis to provide data to inform novel antimicrobial strategies for sepsis in Malawi and similar high-HIV high-TB prevalence settings throughout sSA.

### 1.4.1 Specific aims

The specific aims of this thesis are:

1. To describe the presentation, aetiology, outcome, and determinants of mortality from sepsis in adults presenting to Queen Elizabeth Central Hospital, Blantyre Malawi;
2. To describe the acquisition and carriage of ESBL-E in sepsis survivors, with an analysis of determinants of carriage.

## 1.5 Thesis overview

This thesis is based around a clinical study of sepsis in Blantyre, Malawi, which is described in Chapter 2. Chapter 3 presents data on the clinical presentation, aetiology and outcomes of sepsis in Blantyre, Malawi, with extended exploratory modelling of outcome in Chapter 4. Chapter 5 follows sepsis survivors out of the hospital and presents longitudinal stool sampling data to quantify ESBL-E carriage. Antibiotic-unexposed hospital controls and community members provide comparator cohorts to the antibacterial-exposed sepsis cohorts. To track bacteria and AMR-containing mobile genetic elements within study participants, I have used whole-genome sequencing (WGS) of cultured isolates. Chapter 6 presents the WGS data to give an overview of the genomic landscape of ESBL *E. coli* in Blantyre, whilst Chapter 7 outlines my attempts to use whole genome sequencing as a high-resolution typing tool to track AMR. Chapter 8 develops and fits longitudinal Markov models to understand the determinants of ESBL-E carriage in study participants, and brings in the genomic typing from the previous chapter. Finally, chapter 9 provides conclusions, and suggestions of further work.

## 1.6 Appendix

### 1.6.1 Search terms for sepsis literature review

Sepsis *and* ((Angola *or* Benin *or* Botswana *or* Burkina Faso *or* Burundi *or* Cameroon *or* Cape Verde *or* Central African Republic *or* Chad *or* Comoros *or* Republic of the Congo *or* Congo Brazzaville *or* Democratic republic of the Congo *or* Cote d'Ivoire *or* Djibouti *or* Equatorial Guinea *or* Eritrea *or* Ethiopia *or* Gabon *or* The Gambia *or* Ghana *or* Guinea *or* Guinea-Bissau *or* Kenya *or* Lesotho *or* Liberia *or* Madagascar *or* Malawi *or* Mali *or* Mauritania *or* Mauritius *or* Mozambique *or* Namibia *or* Niger *or* Nigeria *or* Reunion *or* Rwanda *or* Sao Tome and Principe *or* Senegal *or* Seychelles *or* Sierra Leone *or* Somalia *or* South Africa *or* Sudan *or* Swaziland *or* Eswatini *or* Tanzania *or* Togo *or* Uganda *or* Western Sahara *or* Zambia *or* Zimbabwe) *or* Africa).

### 1.6.2 Search terms for ESBL literature review

((ESBL) **or** Extended-spectrum beta-lactamase)) *and* (((Angola *or* Benin *or* Botswana *or* Burkina Faso *or* Burundi *or* Cameroon *or* Cape Verde *or* Central African Republic *or* Chad *or* Comoros *or* Republic of the Congo *or* Congo Brazzaville *or* Democratic republic of the Congo *or* Cote d'Ivoire *or* Djibouti *or* Equatorial Guinea *or* Eritrea *or* Ethiopia *or* Gabon *or* The Gambia *or* Ghana *or* Guinea *or* Guinea-Bissau *or* Kenya *or* Lesotho *or* Liberia *or* Madagascar *or* Malawi *or* Mali *or* Mauritania *or* Mauritius *or* Mozambique *or* Namibia *or* Niger *or* Nigeria *or* Reunion *or* Rwanda *or* Sao Tome and Principe *or* Senegal *or* Seychelles *or* Sierra Leone *or* Somalia *or* South Africa *or* Sudan *or* Swaziland *or* Eswatini *or* Tanzania *or* Togo *or* Uganda *or* Western Sahara *or* Zambia *or* Zimbabwe) *or* Africa)).

Table 1.8: Sepsis diagnostic criteria

Definition	Diagnosis	Criteria
Sepsis-1 (1991)	SIRS	Two or more of: Temperature $> 38^{\circ}\text{C}$ or $< 36^{\circ}\text{C}$ , Heart rate $> 90$ /min, Respiratory rate $> 20$ /min or $\text{PaCO}_2 < 32\text{mmHg}$ (4.3 kPa), White blood cell count $> 12 \times 10^9$ /L or $< 4 \times 10^9$ /L or $> 10\%$ immature forms
	Sepsis	SIRS plus proven or suspected infection
	Severe Sepsis	Sepsis plus acute organ dysfunction
	Septic shock	Sepsis with persistent hypotension after fluid resuscitation
Sepsis-2 (2001)	Sepsis	Infection documented or suspected and some of the following General variables: temperature $> 38^{\circ}\text{C}$ or $< 36^{\circ}\text{C}$ , heart rate $> 90$ min $^{-1}$ or $> 2$ SD above normal for age, tachypnoea, altered mental status, significant oedema or positive fluid balance ( $> 20\text{ml/kg}$ over 24hrs), hyperglycaemia $> 7.7\text{mmol/L}$ Inflammatory variables: white blood cell count $> 12 \times 10^9$ /L or $< 4 \times 10^9$ /L or $> 10\%$ immature forms, plasma C-reactive protein $> 2$ SD above normal, plasma procalcitonin $> 2$ SD above normal Haemodynamic variables: arterial hypotension (SBP $< 90$ mmHg or MAP $< 70$ mmHg or SBP decrease $> 40\text{mmHg}$ in adults or 2SD below normal range, SvO $_2 > 70\%$ , Cardiac index $> 3.5$
	Severe sepsis	Sepsis plus organ dysfunction Organ dysfunction variables: arterial hypoxaemia ( $\text{PaO}_2 / \text{FiO}_2$ ) $< 300$ , acute oliguria (urine output $< 0.5$ ml kg $^{-1}$ hr $^{-1}$ for at least 2 hours), creatinine increase $> 0.5\text{mg/dL}$ , coagulation abnormalities (INR $> 1.5$ or aPTT $> 60\text{s}$ ), ileus, thrombocytopenia (platelet count $< 100,000$ /mL, hyperbilirubinaemia (plasma bilirubin $> 4\text{mg/dL}$ or $70$ mmol /L
	Septic shock	Sepsis plus hypotension SBP $< 90\text{mmHg}$ or MAP $< 60\text{mmHg}$ or reduction in SBP of $40\text{mmHg}$ from baseline despite adequate volume resuscitation
Sepsis-3 (2016)	Sepsis	Infection plus life threatening organ dysfunction defined by an acute change in SOFA score of 2 or more
	Septic shock	Persisting hypotension requiring vasopressors to maintain MAP $65\text{mmHg}$ AND serum lactate below $2\text{mmol/L}$

*Note:*

SIRS = Systemic Inflammatory Response Syndrome, SD = Standard deviation, SBP = Systolic blood pressure, MAP = Mean arterial blood pressure

Table 1.9: Sequential organ failure assessment (SOFA) score

System	Score				
	0	1	2	3	4
<b>Respiratory</b>					
Pao <sub>2</sub> / FiO <sub>2</sub> mmHg (kPa)	400 (53.3)	< 400 (53.3)	< 300 (40)	< 200 (26.7) with respiratory support	< 100 (13.3) with respiratory support
<b>Coagulation</b>					
Platelets x100,000/mL	150	< 150	< 100	< 50	< 20
<b>Liver</b>					
Bilirubin mg /dL (mmol/ L)	<1.2 (20)	1.2-1.9 (20 – 32)	2.0 – 5.9 (33-101)	6.0 – 11.9 (102 – 204)	> 12.0 (204)
<b>Cardiovascular</b>					
Cardiovascular	MAP > 70mmHg	MAP < 70mmHg	Dopamine < 5 or dobutamine any dose	Dopamine 5.1 – 15 or epinephrine < 0.1 or norepinephrine < 0.1	Dopamine > 15 or epinephrine > 0.1 or norepinephrine > 0.1
<b>CNS</b>					
Glasgow coma scale	15	13-14	10-12	7-9	< 6
<b>Renal</b>					
Creatinine mg/dL (mmol /L)	< 1.2 (110)	1.2 – 1.9 (110 – 170)	2.0 – 3.4 (171 – 299)	3.5 – 4.9 (300 – 440)	> 5.9 (440)
Urine output (ml /day)				< 500	< 200

*Note:*

PaO<sub>2</sub> = Arterial partial pressure of oxygen, FiO<sub>2</sub> = Inspired fraction of oxygen, MAP = mean arterial blood pressure, CNS = Central nervous system. All doses of inotropes are micrograms/kg/min

Table 1.10: Selected causes of fever in sSA since 2013

Study	Year	Country	Setting	Patient Population	Test used	Case definition	Confirmed acute disease
<b>Leptospirosis</b>							
Zida 2018	2014-15	Burkina Faso	Central reference lab	Febrile Jaundice adults and children	In house IgM followed by MAT and PCR (acute only, > 1:400)	MAT > 1:400	27/781 (3.5%)
Guillebaud 2018	2014-2015	Madagascar	21 health-care centres	Febrile adults and children	PCR array	Positive PCR	1/682 (0.2%)
Maze 2018	2012-2014	Tanzania	2 Referral Hospitals	Febrile adults and children	MAT (acute + conv)	MAT > 1:800 or fourfold rise	24/1239 (1.9%)
Gadia 2017		Central African Republic	Central reference lab	Febrile Jaundice adults and children	IgM ELISA (acute only)	Any IgM positive	0/198 (0%)
Hagen 2017	2011-2013	Madagascar	District Hospital	Adults and children FUO	PCR	Positive PCR	0/1009 (0%)
Biscornet 2017	2014-2015	Seychelles	Reference leptospirosis clinic	13 or above FUO, referred to central leptospirosis clinic	In house IgM followed by MAT and PCR (acute + conv)	MAT > 1:400 or fourfold rise	51/225 (23%)
Dreyfus 2017	2014	Uganda	2 Health centres	Any adult heath centre attendee	MAT (acute only)	MAT > 1:800	7/359 (1.9%)
Hercik 2017	2014-2015	Tanzania	District hospital	Febrile adults and children	Taqman PCR array	Positive PCR	22/842 (2.6%)
Chipwaza 2015	2014	Tanzania	District hospital	Outpatient febrile children	IgM IgG ELISA then MAT (acute only)	MAT > 1:160	26/200 (13%)



Table 1.10: Selected causes of fever in sSA since 2013 (*continued*)

Study	Year	Country	Setting	Patient Population	Test used	Case definition	Confirmed acute disease
<b>Q-fever</b>							
Amoako 2018	2016-17	Ghana	2 district hospitals	Febrile children	Taqman PCR array	Positive PCR	1/166 (0.6%)
Hercik 2017	2014-2015	Tanzania	District hospital	Febrile adults and children	Taqman PCR array	Positive PCR	2/842 (0.2%)
Boone 2017	2011-13	Madagascar	Two public health care facilities	Febrile adults and children	PCR	Positive PCR	0/1005
Njeru 2016	2014-15	Kenya	Two district hospitals	Febrile adults and children	Phase I/II IgG ELISA and IFA; PCR on subset (acute only)	Phase II IgG IFA titre > 1:128	163/1067 (15%), 10/448 (2.2%) PCR positive
Mourembou 2016	2013-14	Gabon	Four health centres	Febrile children	PCR	Positive PCR	0/410 (0%)
Maina 2016	2011-12	Kenya	District Hospital	Febrile children	IgM/IgG ELISA phase I and II (acute and conv)	Phase II IgG seroconversion	25/370 (8.9%)
Angelaksis 2014	2010-12	Senegal, Mali, Gabon	Six health centres	Febrile adults and children	PCR	Positive PCR	6/1388 (0.4%)
<b>Brucellosis</b>							
Cash-Goldwasser 2018	2012-14	Tanzania	Two referral hospitals	Febrile adults and children	MAT and blood culture (acute + conv)	Fourfold rise in MAT	39/562 (6.9%)

Table 1.10: Selected causes of fever in sSA since 2013 (*continued*)

Study	Year	Country	Setting	Patient Population	Test used	Case definition	Confirmed acute disease
Gafritia 2017	2014	Rwanda	District hospital	Adults, fever	Rose Bengal test	Positive test	10/198 (6.1%)
Boone 2017	2011-13	Madagascar	Two public health care facilities	Febrile adults and children	PCR	Positive PCR	15/1005 (1.5%)
De Glanville 2017	2012	Kenya	Referral hospital and private clinic	Febrile adults and children	Rose Bengal test	Positive test	8/825 (9.7%)
Njeru 2016	2014-15	Kenya	Two district hospitals	Febrile adults and children	Rose bengal test, IgG/IgM ELISA, PCR (acute only)	Positive ELISA or PCR	146/1067 (13.7%)
Chipwaza 2015	2014	Tanzania	District hospital	Outpatient febrile children	IgM and IgG and tube agglutination (acute only)	Positive IgM	26/370 (7.0%)
Feleke 2015	2011	Ethiopia	Health centre	Febrile adults and children	Brucella antigen test	Positive test	3/280 (1%)
<b>Rickettsioses</b>							
Amoako 2018	2016-17	Ghana	2 district hospitals	Febrile children	Taqman PCR array	Positive PCR	5/166 (3.0%) RS
Hercik 2017	2014-2015	Tanzania	District hospital	Febrile adults and children	Taqman PCR array	Positive PCR	2/842 (0.2%) RF
Sothmann 2017	2012	Ghana	Referral hospital	Febrile Children	PCR	Positive PCR	6/431 (1.4%) RF

Table 1.10: Selected causes of fever in sSA since 2013 (*continued*)

Study	Year	Country	Setting	Patient Population	Test used	Case definition	Confirmed acute disease
Maina 2016	2011-12	Kenya	District Hospital	Febrile children	IgG ELISA (acute and conv)	Fourfold rise in IgG titre	63/364 (22.4%) SFG 3/364 (1.1%) TG, 10/364 (3.6%) STG
Elfving 2016	2011	Zanzibar	District hospital	Febrile children with no diagnosis	PCR	Positive PCR	0/83 RS
Mourembou 2015	2013-14	Gabon	4 health centres	Febrile children	PCR	Positive PCR	42/410 (10.2%) RF
<b>Dengue</b>							
Amoako 2018	2016-17	Ghana	2 district hospitals	Febrile children	Taqman PCR array	Positive PCR	2/166 (1.2%)
Guillebaud 2018	2014-2015	Madagascar	21 health-care centres	Febrile adults and children	PCR macroarray	Positive PCR	0/682 (0%)
Kayiwa 2018	2014-2017	Uganda	District hospital	Febrile adults and children	PCR	Positive PCR	1/384 (0.26%)
Makiala-Mandanda 2018	2003-2012	Democratic Republic of Congo	Central lab	Febrile Jaundice, yellow fever IgM negative	PCR	Positive PCR	16/453 (3.5%)
Muianga 2018	2014	Mozambique	Not clear	Febrile adults and children	IgG, IgM and PCR (acute only)	Positive PCR	37/99 by PCR (37.4%)
Mugabe 2018	2016	Mozambique	Five health centres	Febrile adults and children	IgM, IgG, PCR (acute only)	Positive PCR	PCR 0/163
Hercik 2018	2014-2015	Tanzania	District hospital	Febrile adults and children	Taqman PCR array	Positive PCR	1/191 (0.5%)

Table 1.10: Selected causes of fever in sSA since 2013 (*continued*)

Study	Year	Country	Setting	Patient Population	Test used	Case definition	Confirmed acute disease
Gadia 2017		Central African Republic	Central reference lab	Febrile Jaundice adults and children	IgM (Acute only)	Positive IgM	0/198 (0%)
Vu 2017	2014-2015	Kenya	Two health centres	Febrile children	PCR	Positive PCR	82/1104 (7.4%)
Waggoner 2017	2014-2015	Kenya	Two health centres and two district hospitals	Children with fever	PCR	Positive PCR	0/385 (0%)
Kolawole 2017	2016	Nigeria	Two health centres	Adults and children with fever	IgM, IgG, PCR (Acute only)	Positive PCR	11/176 (6.2%)
Nasir 2017	2016	Nigeria	Teaching hospital	Adults and children with fever	NS1 antigen	Positive antigen	15/171 (8.8%)
Ngoi 2016	2014-2015	Kenya	Five health clinics, one district hospital	Adults with fever, negative for acute HIV and malaria	PCR	Positive PCR	43/489 (8.8%)
Onoja 2016	2014	Nigeria	One district hospital	Adults and children with fever	IgM (Acute only)	Positive IgM	64/274 (23.3%)
Kajeguka 2016	2013-2014	Tanzania	Three district hospitals	Probable Dengue (on clinical and IgM)	PCR	Positive PCR	0/381 (0%)
Elfving 2016	2011	Zanzibar	District hospital	Febrile children with no diagnosis	PCR	Positive PCR	0/83
Sow 2016	2009-2013	Senegal	Seven health-care facilities	Adults and children with fever	IgM, PCR (acute only)	Positive PCR	3/13,845 (0.02%)

Table 1.10: Selected causes of fever in sSA since 2013 (*continued*)

Study	Year	Country	Setting	Patient Population	Test used	Case definition	Confirmed acute disease
Chipwaza 2014	2013	Tanzania	One district hospital	Children with fever	IgM, PCR (acute only)	Positive PCR	29/364 (8.0%)
<b>Chikingunya</b>							
Kayiwa 2018	2014-2017	Uganda	District hospital	Febrile adults and children	PCR	Positive PCR	19/384 (4.9%)
Makiala-Mandanda 2018	2003-2012	Democratic Republic of Congo	Central lab	Febrile Jaundice, yellow fever IgM negative	PCR	Positive PCR	2/453 (0.4%)
Muianga 2018	2014	Mozambique	Not clear	Febrile adults and children	IgG, IgM (acute only)	Positive IgM	8/114 by IgM (7%)
Antonio 2018	2015-16	Mozambique	Eight health centres	Undifferentiated fever	IgM, IgG (Acute only)	Positive IgM	6/392 (1.5%)
Mugabe 2018	2016	Mozambique	Five health centres	Febrile adults and children	IgM, IgG, PCR (Acute only)	Positive PCR	PCR 0/163, IgM 17/163 (10.4%)
Sow 2017	2009-2010	Senegal	Fiver health centres and four schools	Febrile adults and children	IgM, IgG, PCR (Acute only)	Positive PCR	20/1049 (1.4%)
Gadia 2017		Central African Republic	Central reference lab	Febrile Jaundice adults and children	IgM (Acute only)	Positive IgM	0/198 (0%)
Olajiga 2017	2015-2016	Nigeria	Seven hospitals	Fever or joint pain or rash, over 10 years	IgM, IgG (acute only)	Positive IgM	66/172 (38.4) by IgM
Waggoner 2017	2014-2015	Kenya	Two health centres and two district hospitals	Children with fever	PCR	Positive PCR	32/385 (8.3%)

Table 1.10: Selected causes of fever in sSA since 2013 (*continued*)

Study	Year	Country	Setting	Patient Population	Test used	Case definition	Confirmed acute disease
Ngoi 2016	2014-2015	Kenya	Five health clinics, one district hospital	Adults with fever, negative for acute HIV and malaria	PCR	Positive PCR	0/489 (0%)
Kajeguka 2016	2013-2014	Tanzania	Three district hospitals	Probable Chikungunya (on clinical and IgM)	PCR	Positive PCR	11/263 (4.2%)
Elfving 2016	2011	Zanzibar	District hospital	Febrile children with no diagnosis	PCR	Positive PCR	0/83
Sow 2016	2009-2013	Senegal	Seven health-care facilities	Adults and children with fever	IgM, PCR (acute only)	Positive PCR	13/13,845 (0.1%)
Chipwaza 2014	2013	Tanzania	One district hospital	Children with fever	IgM (acute only)	Positive IgM	17/364 (4.7%)
<b>Zika</b>							
Kayiwa 2018	2014-2017	Uganda	District hospital	Febrile adults and children	PCR	Positive PCR	5/384 (1.3%)
Makiala-Mandanda 2018	2003-2012	Democratic Republic of Congo	Central lab	Febrile Jaundice, yellow fever IgM negative	PCR	Positive PCR	0/453 (0%)
Sow 2016	2009-2013	Senegal	Seven health-care facilities	Adults and children with fever	IgM, PCR (Acute only)	Positive PCR	9/13,845 (0.1%)

*Note:*

RS = Rickettsia spp., RF = R. felis, SFG/TG/STG = spotted fever/ typhus/scrub typhus group

Table 1.11: included studies providing an estimate of proportion of ESBL producers in invasive *E. coli* and *K. pneumoniae* isolates in sSA.

Year	First author	Country	Population	Sample	<i>E. coli</i>	<i>K. pneumoniae</i>
2018	Guiral	Mozambique	A C IP	Blood urine	13/151 (9%)	ND
2018	Karppinen	Angola	C IP OP	Wound swab	8/15 (53%)	10/13 (77%)
2018	Kpoda	Burkina Faso	A C IP OP	Various	117/296 (40%)	48/109 (44%)
2018	Onanuga	Nigeria	A OP	Urine	4/18 (22%)	30/45 (67%)
2018	Seni	Nigeria	A C IP OP	Various	41/60 (68%)	ND
2018	Zeynudin	Ethiopia	NA IP OP	Various	13/13 (100%)	30/31 (97%)
2017	Ampaire	Uganda	A C IP OP	Various	18/146 (12%)	10/68 (15%)
2017	Andrew	Uganda	A C IP OP	Various	33/44 (75%)	33/36 (92%)
2017	Archary	South Africa	C IP	Various	2/11 (18%)	13/19 (68%)
2017	Henson	Kenya	A C IP OP	Blood	ND	101/198 (51%)
2017	Ibrahim	Nigeria	A C IP OP	Urine wound swab	68/140 (49%)	76/108 (70%)
2017	Kassam	Tanzania	A C IP OP	Wound swab	6/14 (43%)	8/11 (73%)
2017	Legese	Ethiopia	C IP	Blood urine	5/6 (83%)	16/19 (84%)
2017	Manyahi	Tanzania	A C IP OP	Urine	15/110 (14%)	9/27 (33%)
2017	Sangare	Mali	A C IP	Blood	20/31 (65%)	20/26 (77%)
2017	Vasaikar	South Africa	A C IP OP	Various	ND	117/169 (69%)
2016	Abera	Ethiopia	A C IP OP	Blood urine	71/122 (58%)	34/49 (69%)
2016	Agyekum	Ghana	A C IP OP	Blood urine	30/58 (52%)	33/43 (77%)
2016	Breurec	Senegal	C IP	Blood CSF	ND	33/41 (80%)
2016	Buys	South Africa	C IP	Blood	ND	339/410 (83%)
2016	Eibach	Ghana	A C IP	Blood	5/50 (10%)	34/41 (83%)
2016	Kabwe	Zambia	C IP	Blood	5/5 (100%)	71/74 (96%)
2016	Leski	Sierra Leone	A C OP	Urine	0/13 (0%)	9/15 (60%)
2016	Mohammed	Nigeria	A C IP OP	Various	41/172 (24%)	59/178 (33%)
2016	Naas	Madagascar	C IP OP	Blood	0/7 (0%)	11/14 (79%)
2016	Ndir	Senegal	C IP	Blood	7/12 (58%)	33/40 (82%)
2016	Ouedraogo	Burkina Faso	A C IP OP	Various	121/202 (60%)	46/70 (66%)
2016	Sangare	Mali	A C IP	Blood	8/11 (73%)	10/14 (71%)
2016	Seni	Tanzania	A IP	Pertitoneal fluid	7/19 (37%)	5/10 (50%)
2015	Dramowski	South Africa	C IP OP	Blood	14/97 (14%)	119/154 (77%)

Table 1.11: included studies providing an estimate of proportion of ESBL producers in invasive *E. coli* and *K. pneumoniae* isolates in sSA. (continued)

Year	First author	Country	Population	Sample	<i>E. coli</i>	<i>K. pneumoniae</i>
2015	Irenge	Democratic Republic of Congo	A C IP OP	Blood	9/54 (17%)	10/21 (48%)
2015	Kateregga	Uganda	A C IP OP	Various	36/64 (56%)	24/33 (73%)
2015	Opintan	Ghana	A C IP OP	Various	81/440 (18%)	ND
2015	Pons	Mozambique	A C IP OP	Blood urine	ND	16/50 (32%)
2015	Rafa	Central african republic	A C IP	Wound swab	33/47 (70%)	10/19 (53%)
2014	Adeyankinnu	Nigeria	A C IP OP	Various	36/135 (27%)	16/62 (26%)
2014	Irenge	Democratic Republic of Congo	A C IP OP	Urine	57/376 (15%)	ND
2014	Scherbaum	Gabon	A IP	Various	5/14 (36%)	3/6 (50%)
2014	Yusuf	Nigeria	A IP OP	Various	47/278 (17%)	19/128 (15%)
2013	Alabi	Gabon	A C IP OP	Various	ND	43/85 (51%)
2013	Ibrahim	Sudan	A C IP OP	Various	70/232 (30%)	ND
2013	Obeng-Nkrumah	Ghana	A C IP OP	Various	55/126 (44%)	59/96 (61%)
2013	Raji	Nigeria	A C IP OP	Various	21/43 (49%)	12/32 (38%)
2013	van der Meeren	Mozambique	C IP	Urine	9/14 (64%)	15/17 (88%)
2011	Idowu	Nigeria	A IP	Wound swab	6/15 (40%)	ND
2010	Moyo	Tanzania	A C IP OP	Urine	54/138 (39%)	ND
2009	Bercion	Central African Republic	A C OP	Urine	29/357 (8%)	17/57 (30%)
2009	Mshana	Tanzania	A C IP OP	Various	31/127 (24%)	58/91 (64%)
2007	Sire	Senegal	A C IP OP	Urine	38/1010 (4%)	ND
2005	Blomberg	Tanzania	C IP OP	Blood	9/36 (25%)	9/48 (19%)
2005	Gangoue Pieboji	Cameroon	A C IP	Various	13/91 (14%)	12/64 (19%)
2005	Ndugulile	Tanzania	A IP	Various	4/13 (31%)	2/2 (100%)
2004	Dromigny	Senegal	A C OP	Urine	2/233 (1%)	1/34 (3%)
2002	Dromigny	Senegal	A C OP	Urine	1/386 (0%)	6/97 (6%)

Note:

A = Adults, C = children, IP = inpatients OP = outpatients, ND = not done.



Table 1.12: Included studies providing estimate of prevalence of ESBL-E gut mucosal colonisation in sSA.

Study	Year Pub.	Study Period	Country	Study Type	Inclusion	Age group	Median age	n
Ruppe	2009	NR	Senegal	Cross sec.	Children in village selected for remoteness	Children	6.9yr*	20
Tande	2009	2003	Mali	Cross sec.	Orphanage children	Children	NR	38
					Orphanage staff	Adults	NR	30
Andriatahina	2010	2008	Madagascar	Cohort	Inpatients	Children	38.3m	244
Herindrainy	2011	2009	Madagascar	Cross sec.	Health centre attendees	Adults	NR	306
					Health centre attendees	Children	NR	147
Woerther	2011	2007-08	Niger	Cohort	Children with SAM, inpatients	Children	16.3m*	55
Albrechtova	2012	2009	Kenya	Cross sec.	Community members	Adults	NR	23
Isendahl	2012	2010	Guinea-Bissau	Cross sec.	Children att. hospital w/ fever or tachycardia	Children	NR	408
Lonchel	2012	2009	Cameroon	Cross sec.	Students in the community	Adults	24.7yr*	150
					Outpatients	Adults	36.9yr*	208
Lonchel	2013	2009	Cameroon	Cross sec.	Inpatients	Adults	46.8yr*	121
Magoue	2013	2010	Cameroon	Cross sec.	Hospital workers and their families	Adults	NR	87
					Inpatients	Adults	NR	208
					Relatives and carers of inpatients	Adults	NR	63
					Outpatients	Adults	NR	232
Schaumburg	2013	2010-11	Gabon	Cross sec.	Hospital inpatients	Children	NR	200
Nelson	2014	2013	Tanzania	Cohort	Pregnant women and neonates, inpatient	Neonate	0d	126
						Adults	26.5yr*	113
Chereau	2015	2013-14	Madagascar	Cross sec.	Pregnant women in the community	Adults	26yr*	356
Desta	2016	2012	Ethiopia	Cross sec.	Inpatients	Adults	35yr	154
					Inpatients	Children	7yr	94

Table 1.12: Included studies providing estimate of prevalence of ESBL-E gut mucosal colonisation in sSA. (*continued*)

Study	Year Pub.	Study Period	Country	Study Type	Inclusion	Age group	Median age	n
					Inpatients	Neonate	9d	19
Djuikoue	2016	2011-12	Cameroon	Cross sec.	Outpatient women with susp. UTI	Adults	NR	86
Farra	2016	2013	CAR	Cross sec.	Healthy community controls from diarrhoea study	Children	10.5m	134
Kurz	2016	2014	Rwanda	Cohort	Inpatients and one main caregiver	both	29yr	753
Mshana	2016	2014	Tanzania	Cross sec.	Community members	both	10yr	334
Ribeiro	2016	2013	Angola	Cross sec.	Community members no antibiotics/hospital exposure last 3 m	Adults	NR	18
Tellevik	2016	2010-11	Tanzania	Cross sec.	<2yr attending health centre for vaccine	Children	NR	250
					Inpatients	Children	NR	353
Magwenzi	2017	2015	Zimbabwe	Cohort	Inpatient within 24hr of admission	Children	1.0yr	164
Moremi	2017	2015	Tanzania	Cross sec.	Street children	Children	14.2yr*	107
Wilmore	2017	2014-15	Zimbabwe	Cross sec.	Outpatient, HIV infected, stable on ART	Children	11yr	175
Chirindze	2018	2016	Mozambique	Cross sec.	Students in the community	Adults	NR	275
Founou	2018	2017	South Africa	Cohort	On hospital admission	Adults	NR	43
Herindrainy	2018	2015-16	Madagascar	Cross sec.	Pregnant women at delivery (home/facility)	Adults	26yr*	275
Katakweba	2018	2011-13	Tanzania	Cross sec.	Community members	Adults	NR	70
Marando	2018	2016	Tanzania	Cross sec.	Neonates with sepsis	Neonate	6d	304
Moremi	2018	2014-15	Tanzania	Cohort	On hospital admission	Adults	NR	930
Nikema	2018	2015-16	Togo	Cross sec.	<5yr with febrile gastroenteritis	Children	NR	81
Sanneh	2018	2015	The Gambia	Cross sec.	Food handlers in schools	Adults	37yr*	565

Table 1.12: Included studies providing estimate of prevalence of ESBL-E gut mucosal colonisation in sSA. (*continued*)

Study	Year Pub.	Study Period	Country	Study Type	Inclusion	Age group	Median age	n
Stanley	2018	2017	Uganda	Cross sec.	Participants who reared animals, attending health facility with a fever and/or diarrhoea but without malaria	both	21.7yr*	300

*Note:*

NR = Not reported, CAR = Central African Republic.

\* Mean not median



# Chapter 2

## Methods

### 2.1 Chapter Overview

This chapter gives an overview of the clinical study which underpins this thesis, and the laboratory procedures used in analysis. Computational and statistical procedures are largely detailed in the relevant chapters as are further specific details, where necessary.

### 2.2 Study site

#### 2.2.1 Malawi

Malawi is a country of 17.5 million people in South-Eastern Africa[281]. It is one of the poorest countries in the world: it is a low income country under the World Bank classification, with a 2017 Gross National Income (GNI) per capita of \$320 in US dollars[282], and was ranked 171st of 189 countries in 2017 by the human development index (HDI), a composite statistic of life expectancy, education and income per capita indicators[283]. In 2010, 71% of the population was estimated to survive on less than \$1.90 per day. Life expectancy at birth in 2017 was 63 years, and though significant progress is being made, neonatal and under-5 mortality remains high at 23 and 55 per 1000 live births, respectively. The population is largely rural (83% in 2017), with a young population (44% under the age of 15, 2017) and high fertility rate[282]. Malaria is endemic, and there is an ongoing generalised HIV epidemic: adult HIV prevalence (age 15-49) was estimated to be 9.6% in 2017, though falling from a peak of 16.6% in 1999[284]. HIV antiretroviral therapy (ART) national scale up began in 2004 and in 2017 71% of eligible adults and children were estimated to be receiving ART[284]. It is classed by the WHO as a high-TB/high-HIV burden country, with an estimated TB incidence rate of 131 (95% CI 70-210) cases per 100,000 population



Figure 2.1: Malawi, showing administrative boundaries (North, Central, and South regions), Lilongwe, the capital city and Blantyre, the study location. Source: openstreetmap.org, used under Creative Commons Attribution ShareAlike 2.0 licence CC-BY-SA

per year[285].

It has a subtropical climate, with three main seasons: a warm wet season from November to April, a cooler dry winter period from May to August and a hot dry period from September to October. Blantyre city, the location of the study in this thesis, is the second city of Malawi with a population of 585,000. It is located in Blantyre district, population 995,000 in 2018[281], in the Shire highlands at an altitude of 1000m (Figure 2.1).

### 2.2.2 Queen Elizabeth Central Hospital

Queen Elizabeth Central Hospital (QECH), located in Blantyre city, is the tertiary referral hospital for the Southern Region of Malawi. It has 1300 beds but often operates above capacity, and is the only site providing free inpatient healthcare to the adult population of Blantyre district. Since 2011 it has had a dedicated emergency department for adults, the Adult Emergency and Trauma Centre (AETC), staffed 24 hours a day. Since 2015 (and for the whole of the study period), attendees to the AETC must be referred from a

primary health clinic. Adults attending the AETC are triaged by a nurse and then reviewed by a doctor or clinical officer; if admission under a speciality team (including medicine) is deemed appropriate then a patient will be reviewed by an intern or registrar from the admitting speciality and usually by a consultant within 24 hours. There is a 6-bed AETC resuscitation area in which oxygen concentrators, cardiac monitors and a defibrillator are available; none of these items are available in the rest of the AETC.

There are two dedicated single-sex medical wards, each of approximately 60 beds, and one mixed-sex TB ward. Male and female high-dependency units (HDUs), each with a capacity of six beds, have oxygen concentrators (or, if available, oxygen cylinders) to deliver supplemental oxygen. The medical wards are staffed by two or three trained nurses and a variable number of nursing students. Basic nursing care is usually provided by a patients relative or friend, called a ‘guardian.’ Patients on the medical wards are reviewed twice-weekly by a consultant physician and then variably at other times by junior doctors, clinical officers or medical students depending on the availability of staff. Malawi national treatment guidelines suggest ceftriaxone for the treatment of sepsis requiring hospitalisation.

### **2.2.3 Participating Laboratories**

#### **2.2.3.1 Malawi-Liverpool-Wellcome Clinical Research Programme**

The Malawi-Liverpool Wellcome Trust Clinical Research Programme (MLW) was established in 1995 and since then has been active in researching priority health issues in Malawi. It is an affiliate of the Malawi College of Medicine, and is based in the grounds of QECH in Blantyre. It runs an on-site microbiology laboratory which has provided an aerobic blood culture service to QECH since 1998, and also provides cerebrospinal fluid (CSF) microscopy and culture. Bacterial culture is carried out as per British Society of Antimicrobial Chemotherapy (BSAC) guidelines[286] and the laboratory adheres to UK NEQAS external quality control. It is core funded by the Wellcome Trust. MLW operates the Malawi College of Medicine Tuberculosis Laboratory where sputum GeneXpert testing and mycobacterial blood culture were undertaken.

#### **2.2.3.2 Wellcome Trust Sanger Institute**

The Wellcome Sanger Institute (WSI) is a research institute based in Hinxton, UK, which was established in 1993, and undertakes research in all aspects of genomics including bacterial genomics as part of the parasite and microbes programme. It has one of the largest DNA sequencing facilities in the world as well as exceptional high performance computing cluster capacity. It is funded by the Wellcome Trust.

## 2.3 Clinical Study

The DASSIM (Developing an Antimicrobial Strategy for Sepsis in Malawi) study was an observational cohort study, recruiting from the AETC at QECH with two broad aims: firstly, to describe the presentation, aetiology and determinants of outcome in sepsis in Malawi; and secondly to determinants of carriage of ESBL-E in sepsis survivors.

### 2.3.1 Objectives

The study was open for recruitment between 19 February 2017 and 2 October 2018; there were three arms:

1. Adults with sepsis attending AETC, recruited as early as possible in their attendance to QECH following triage.
2. Antibiotic unexposed adults attending AETC with no plan for antimicrobial administration.
3. Antibiotic-unexposed community members.

Arm 1 (sepsis) participants were recruited in order to clearly define the aetiology of sepsis; arm 2 to clearly define the relative effects of antimicrobial exposure versus hospital admission in ESBL-E acquisition and carriage; and arm 3 to define baseline flux of ESBL-E.

### 2.3.2 Recruitment criteria

Detailed inclusion and exclusion criteria for each arm are given shown in Table 2.1. For logistic reasons, recruitment occurred 7am - 5pm Monday to Friday.

Exclusion criteria were the same for all arms of the study: Participants were not eligible for enrolment if and of the following were true:

- They were unable to give informed consent and no guardian was available to provide proxy consent.
- They spoke neither English nor Chichewa.
- They lived  $> 30$ km from Blantyre city.

The antibiotic-unexposed inpatients and community members were matched on age ( $\pm 5$ yr) and sex to sepsis survivors (defined as patients surviving to 28 days). In addition community members were matched on location to sepsis survivors; putative households for recruitment were identified by random walk from the houses of surviving sepsis participants, with initial direction established by spinning a bottle on the floor.



Table 2.1: Study inclusion criteria

Study Arm	Inclusion Criteria
Arm 1 - Sepsis	Adults (16 years or over) attending AETC AND Axillary temperature > 37.5C or history of fever within 72 hours AND Life threatening organ dysfunction defined by any one of: Oxygen saturations < 90 percent on air, respiratory rate > 30 breaths/minute, systolic blood pressure < 90mmHg, glasgow coma scale < 15
Arm 2 - Inpatient	Adults (16 years or over) attending AETC AND No suspicion of infection or plan for antimicrobial administration AND No antimicrobial therapy within last 4 weeks
Arm 3 - Community	Adults (16 years or over) AND No antimicrobial therapy within last 4 weeks

*Note:*

AETC = Adult Emergency and Trauma Centre

### 2.3.3 Study Visits and Patient Sampling

#### 2.3.3.1 Enrolment assessment

An overview of the study visit schedule is shown in Figure 2.2. At enrolment, following informed written consent, a baseline questionnaire administered to the patient (or the guardian if the participant was obtunded) to capture background demographic and clinical data, a Chichewa-language EQ-5D-3L health related quality of life questionnaire completed, and sample collection undertaken (see below). For hospitalised patients, enrolment aimed to be as soon after triage as possible. Data on therapies administered were captured by the study team including the time of administration of antimicrobial therapy and volumes of intravenous fluid administered, the latter visually confirmed by the study team member. Vital signs were measured by trained study nurses. All blood pressure measurements were made non-invasively with an automated cuff (Omron M2, Omron, Japan), oxygen saturations measured with a dedicated study pulse oximeter (Contec CM50, Contec Medical Systems, China), and temperature measured in the axilla with a digital thermometer (Omron FWH000, Omron, Japan).

All treatment decisions were at the discretion of the participant's attending healthcare

worker.

### **2.3.3.2 Subsequent visits**

Hospitalised participants were reviewed daily Monday to Friday following admission whilst on the wards to capture details of therapies administered. The visit schedule for the study is shown in Figure 2.2. Arm 1 (sepsis) and 2 (hospital inpatients) were reviewed by a study team member 7 and 28 days following enrolment, and at 3 and 6 months; community members were reviewed at 28 days and 6 months. At each visit, details of any antimicrobials or other therapies received and any health care facility contact were collected, and the Chichewa language EQ-5D-3L was repeated. Samples were collected as below.

These study visits occurred preferentially at QECH, but at the patient's home if it was difficult for them to attend the hospital. If patients missed a scheduled visit, then attempts were made to contact them by telephone. If these attempts were unsuccessful then attempts to visit them at home were undertaken by a member of the study team.

### **2.3.3.3 Blood, urine, and stool, sputum and CSF collection**

Blood was collected from arm 1 (sepsis) participants only on enrolment (Figure 2.2) and at the 4-week visit. At baseline, blood was collected aseptically directly into collection tubes with a vacutainer device with the following order of draw and volumes: one BacT/Alert (BioMerieux, France) aerobic blood culture bottle (7-10ml), one BD BACTEC Myco/F Lytic mycobacterial culture bottle (3-5ml) (Beckton Dickinson, United States), one serum (10ml) and two EDTA (4ml, both Grenier Bio-One, Austria) samples. Finger prick for capillary blood was used for point of care diagnostics as described below. At the 4-week visit, 10mls of convalescent serum was collected. At baseline, urine was collected from all arm 1 (sepsis) participants into a sterile polypropylene universal container (Alpha Laboratories, UK) either by the participant themselves (if this was possible) or with the aid of a disposable bedpan.

At the enrolment visit and all other visits for all arms of the study stool was collected into a sterile polypropylene universal container (Alpha Laboratories, UK) with aseptic technique. If it was not possible for a patient to provide stool then a rectal swab was taken using a sterile rayon-tipped swab (Technical Service Consultants Ltd, UK) inserted into 2-3cm into the rectum, rotated for approximately 10 seconds and placed into Amies gel media for transport to the laboratory.

The decision to collect sputum for Xpert testing for tuberculosis or to perform lumbar puncture (LP) rested with the participant's attending healthcare worker. Lumbar puncture, where done, was carried out by QECH AETC or medicine department doctors or clinical

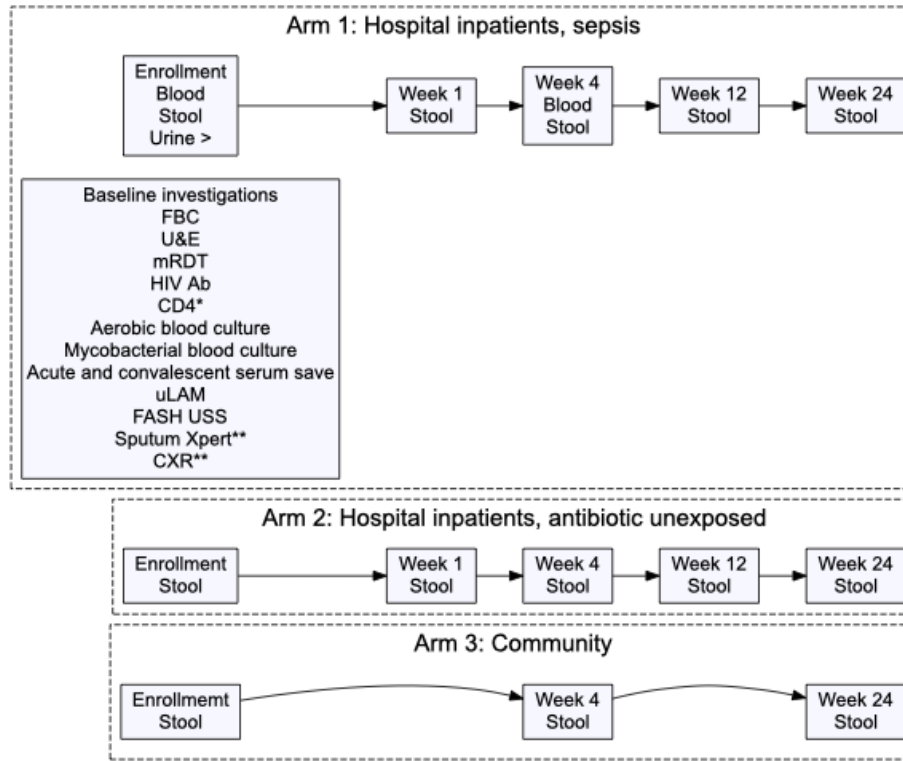


Figure 2.2: Overview of patient sampling schedule. FBC = Full blood count, U&E = Urea and electrolytes, mRDT = Malaria rapid diagnostic test, uLAM = Urinary LAM, FASH USS = Focussed assessment for HIV-associated tuberculosis ultrasound scan

officers and 7-10ml of cerebrospinal fluid (CSF) was aseptically collected. Sputum samples were collected in sterile polypropylene universal container and transported to the Malawi College of Medicine (CoM) tuberculosis laboratory. Myco/F lytic bottles for mycobacterial blood culture were also transported to the CoM tuberculosis laboratory; all other samples were transported to the MLW laboratories in the first instance.

### 2.3.4 Outcomes and sample size calculations

The two co-primary outcomes for the study are:

1. 28-day mortality.
2. Acquisition of ESBL-E detected in stool by aerobic culture at 28 days.

Two power calculations were undertaken: to inform the size of the sepsis cohort, and one for the non-sepsis participants. The sepsis component was powered to detect clinically relevant risk factors for death; we made the a priori assumption that a risk ratio of 2 or more is clinically relevant in this setting and patient population. The initial calculation

suggested 250 patients with severe sepsis with 50% mortality (as seen in previous studies in Malawi) would have 80% power to detect risk factors for death with prevalence of 20-50% that confer a risk ratio (RR) of 1.5-3; therefore, a RR of 2 is likely to be detected. Logistic considerations resulted in a reduction of a feasible sample size to 225; repeating a power calculation with this number of participants suggested 80% power to detect risk factors conferring a RR of 2 with 25-50% prevalence, assuming 50% mortality.

For the second co-primary outcome, assuming 30% of participants with sepsis and hence antimicrobial exposure acquire ESBL-E by 28 days, and that 50% of them die by day 28, 125 antibiotic unexposed hospitalised participants would give 80% power to detect a 50% relative difference in acquisition. Logistic considerations resulted in the target sample size being reduced to 100 participants; revised power calculation suggested that the 80% power to detect a 50% relative difference was maintained.

In order to define the baseline flux of ESBL-E organisms and ESBL genes, one community member for each antibiotic unexposed hospitalised participant, giving a target sample size of 100.

## 2.4 Diagnostic Laboratory Procedures

Results of all laboratory diagnostic procedures were fed back to participant's attending healthcare worker in real time. If any investigation result became available after the participant was discharged (e.g. Mycobacteraemia was identified) then they were contacted and called back to the hospital to be reviewed by me, and referred for care to the relevant QECH department.

### 2.4.1 Point of care diagnostics

Point of care tests were carried out on capillary blood for capillary lactate (Lactate Pro 2, Arkay, Japan), for *P. falciparum* HRP-2 antigen (Paracheck Pf, Orchid Biomedical, India) and, if HIV status was unknown, to test for HIV antibodies using Determine HIV 1/2 kit (Abbott Diagnostic Division) and Unigold HIV 1/2 kit (Trinity Biotech Inc.) following Malawian national guidelines[287].

### 2.4.2 Laboratory diagnostics

#### 2.4.2.1 Haematology and biochemistry

Automated full blood counts (Beckman Coulter HmX Haematology Analyser, Beckman Coulter, USA) were undertaken in the MLW laboratories on EDTA whole-blood samples,

as were CD4 cell counts (Becton Dickinson FACSCount, Becton Dickinson, USA) for HIV-infected participants. Serum samples were allowed to settle upright for 30-60 minutes and then centrifuged (at 1300 g for 10 min). Biochemistry testing for urea and electrolytes (Beckman Coulter AU480 Chemistry Analyser, Beckman Coulter, USA) was undertaken immediately and a maximum of three 1.8ml aliquots of serum stored at -80°C for subsequent analysis.

#### 2.4.2.2 Aerobic blood and CSF culture

Blood in aerobic culture bottles was incubated in an automated system (BacT/Alert BioMerieux) and any bottles that flagged as positive were further processed by Gram's stain and subculture with any pathogens identified to genus level (all) and species level in the case of Enterobacteriaceae using the API system (Biomérieux, France) and standard techniques[288]. Anaerobic culture was not available. Coagulase-negative Staphylococci, Bacillus spp., diphtheroids and alpha-haemolytic Streptococci other than *S. pneumoniae* were considered as contaminants. Antimicrobial sensitivity testing was undertaken using the disc diffusion method following BSAC guidelines.

CSF, where available, was cultured onto blood, chocolate and Sabouraud agar and then identification of any growth undertaken as above.

#### 2.4.2.3 Mycobacterial blood culture

Blood in Myco/F Lytic bottles was cultured at 37°C in and inspected daily for the first 14 days with a handheld UV Wood's lamp, and once every two days thereafter. Contents of the bottles were centrifuged (3000xg for 20 minutes) within 48hr of detection of fluorescence, examined by ZN and Gram stain to exclude contamination, then inoculated into mycobacterial growth indicator tubes (MGIT, Becton Dickinson Diagnostic Systems, United States) for up to 6 weeks. Isolates were classified as *M. tuberculosis* or nontuberculous Mycobacteria by microscopic cording and MPT-64 lateral flow assays (TAUNS Laboratories, Japan).

After 8 weeks of culture of Myco/F Lytic bottles if no fluorescence was seen, then centrifugation and microscopy was carried out. If no organisms were seen, the culture was reported as "no growth."

#### 2.4.2.4 Sputum Xpert

Sputum Xpert testing (Cepheid, United States) was carried out as per the manufacturers instructions: sputum specimens were mixed with sample reagent and incubated at room

temperature for 15 minutes. The liquefied specimen was then loaded into the Xpert MTB/RIF test cartridge for processing.

#### 2.4.2.5 Urinary LAM

Urinary LAM testing was carried out on all available urine samples from HIV-infected participants, using the Alere Determine TB LAM Ag lateral flow assay (Alere, United States) following the manufacturer's instructions. Frozen urine samples were used: they were allowed to come to room temperature and then briefly centrifuged to remove debris. 60 microlitres of urine was applied to the sample pad and the result read after 25 minutes by comparing to the provided reference scale card. If a line was visible in the patient window of the test and darker than the lightest positive line on the reference card then it was considered positive. If a line was visible but lighter than the lightest positive line on the reference card, or if no line was visible, then it was considered negative. If no line was visible in the control window of the test then the test was considered invalid, and repeated. The results were read independently by two readers, who were unaware of the other reader's finding. A tie-break read by a third reader who was unaware of the findings of the other two readers was undertaken in the even of disagreement.

#### 2.4.2.6 Selective stool culture for ESBL-E

Stool and rectal swabs received in the laboratory were stored at 4°C pending processing, before being plated onto commercially available ESBL selective media (CHROMagar ESBL media, CHROMagar, France) and cultured aerobically at 37°C overnight. Rectal swabs were streaked directly onto the plate, or a cotton tipped applicator used for solid stool. CHROMagar is also chromagenic: *E. coli* colonies are pink, *Klebsiella* spp., *Enterobacter* spp. and *Citrobacter* spp. blue and other species white (Figure 2.4). Blue or white colonies were speciated using the API 20E system (Biomerieux, France). Morphologically distinct colonies were confirmed to be ESBL producers using combination disc methods: the putative ESBL producer was cultured overnight on ISO-sensitest agar with discs of cefotaxime and ceftazidime (30 micrograms) with and without clavulanic acid (10 micrograms), and ESBL production confirmed if there was a difference of at least 5mm between discs with and without clavulanic acid (Figure 2.5). For organisms likely to carry a chromosomal AmpC and hence be capable of hydrolysing cefotaxime and ceftazadime, (*Enterobacter* spp., *Citrobacter freundii*, *Morganella morganii*, *Providencia stuartii*, *Serratia* spp., *Hafnia alvei*) cefipime (30 micrograms), an AmpC-stable cephalosporin was used with and without clavulanic acid (10 micrograms). Antimicrobial sensitivity testing to meropenem, amikacin, gentamicin, co-trimoxazole, chloramphenicol and ciprofloxacin following BSAC guidelines[286] was undertaken for a subset of confirmed ESBL-producers.

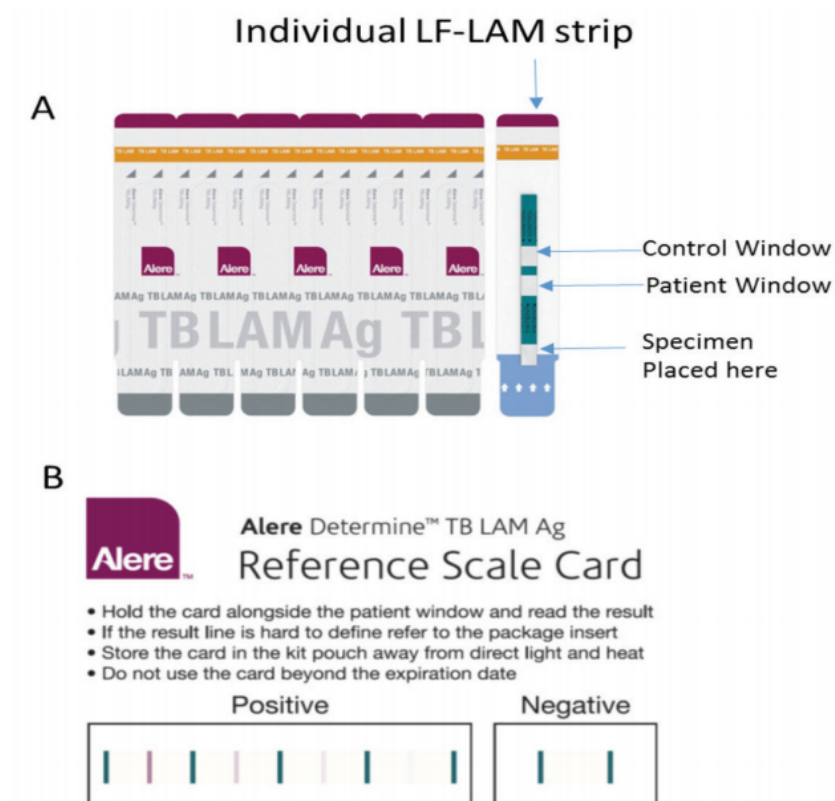


Figure 2.3: A Appearance of urinary LAM tests. B: Urinary LAM reference card. Any line in the patient window darker than the lightest positive line in the reference card is considered positive. Any line lighter than this was classed as negative. Reproduced from [289] .

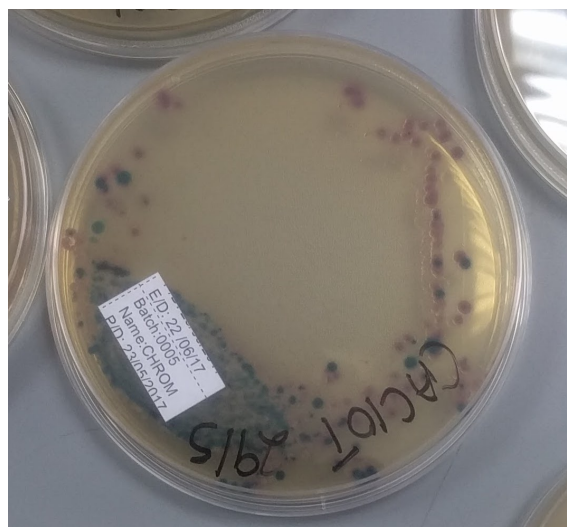


Figure 2.4: Representative example of chromogenic agar appearance showing typical pink colony appearance of *E. coli* and blue colony appearance of *Klebsiella*, *Citrobacter* and *Enterobacter* spp.

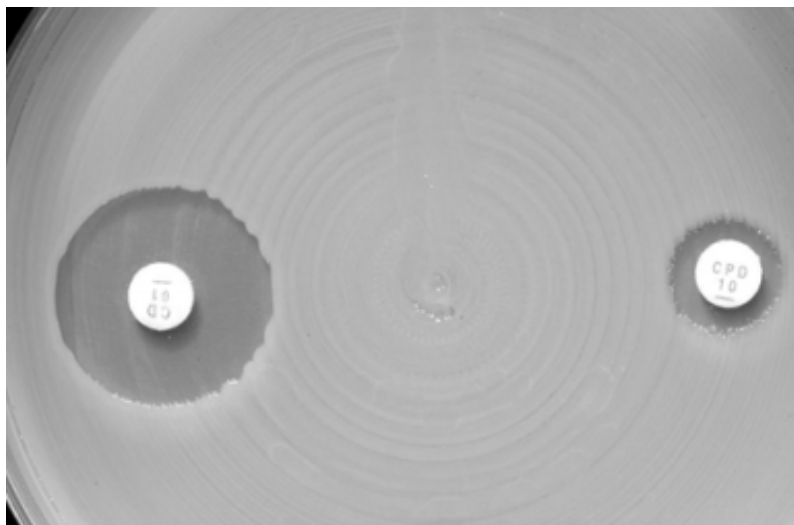


Figure 2.5: Example of combination disc ESBL confirmatory testing. In this example cefpodoxime discs with and without clavulanic acid (a  $\beta$ -lactamase inhibitor) are used. The small zone of inhibition around cefpodoxime (CPD) indicates resistance to cefpodoxime; the larger zone ( $> 5$ mm larger) around the combination cefpodoxime and clavulanic acid disc (CD) confirms this isolate as an ESBL-producer. Reproduced from [290].

### 2.4.3 Case definitions

Bloodstream infection is defined as any pathogenic organism isolated from at least one aerobic blood culture. Coagulase-negative *Staphylococci*, *Bacillus* spp., diphtheroids and alpha-haemolytic *Streptococci* other than *S. pneumoniae* were considered as contaminants. Malaria was defined as a positive malaria rapid test. Meningitis was defined as either a positive CSF culture for a pathogenic organism - excluding the same contaminants as for bloodstream infection - or a positive cryptococcal antigen test on CSF. Tuberculosis was defined as any of positive sputum Xpert, positive urinary LAM at any grade, or positive mycobacterial blood culture.

## 2.5 Molecular methods

One of each morphologically distinct colony from ESBL-E selective culture was taken forward for DNA extraction and whole genome sequencing. DNA was extracted from overnight nutrient broth culture using the Qiagen DNA mini kit (Qiagen, Germany) following the manufacturers instructions. Extracted DNA was shipped to the Wellcome Sanger Institute where it underwent library preparation according to the Illumina protocol and paired-end 150bp sequencing on Illumina HiSeq X10 by the WSI DNA pipelines team. Details of the bioinformatic analyses undertaken are given in the relevant chapters.



## 2.6 Statistical Analysis

Details of analysis methods are given in the relevant chapters. All analyses were undertaken using R v3.6.0 and any Bayesian modelling using Markov Chain Monte Carlo (MCMC) methods using Stan v2.18 via the *rstan* interface with R. This thesis was written using the bookdown package in R and the code to generate it as well as all analysis scripts are available at <https://github.com/joelewis101/thesis>. Unless otherwise states, all boxplots show median and IQR as boxes, 1.5 times IQR as whiskers and outliers (any points outside the whiskers) as points.

## 2.7 Study Team

I am the principal investigator of the study, and led a study team which consisted of: Emma Smith, a medical doctor who managed the running of the study from March 2017 - May 2017; study nurses Lucy Keyala, Grace Mwaminawa and Tusekile Phiri who recruited patients, collected samples and completed follow up visits; field workers Witness Mtambo and Gladys Namacha who recruited patients, collected samples and completed follow up visits; and laboratory technicians Madalitso Mphasa and Rachel Banda who processed laboratory specimens, did stool culture and DNA extractions, and the uLAM testing. The MLW core laboratory staff undertook haematology and biochemistry testing and blood and CSF culture, and the Malawi CoM tuberculosis laboratory staff carried out tuberculosis culture and Xpert testing. Lumbani Makhaza was the data officer in the MLW data team who built and designed and built the study database and electronic and paper data capture forms.

## 2.8 Data Collection and Storage

Data was captured electronically using Open Data Kit software[291] (ODK) and structured TeleForm paper forms (OpenText, Waterloo, Canada). Electronic ODK forms were loaded onto Asos ZenTouch tablets running Android using the ODK Collect Android app. Completed forms were pushed daily to the dedicated secure study SQL (structured query language) database built and administered by the MLW data team. Teleform forms were checked by me, batched and scanned by the MLW data team, and variable values automatically extracted by the TeleForm system; validation was undertaken by the MLW data team and data queries generated for missing or invalid values. Queries were resolved either by manually reconciling with the forms or by discussion with the clinical team. Once the data passed validation it was pushed to the SQL database. Completed paper TeleForm records were stored securely in the MLW data department. All data on the

study database is stored securely with access restricted to the study PI and the study data officer and head of the MLW data team. Other paper records including consent forms were stored in a locked cabinet with access restricted to members of the study team. Results of laboratory investigations in the MLW laboratory were stored in the MLW PreLink laboratory information management system (LIMS), anonymised and linked only to the participant unique study ID number. For analysis, anonymised data were extracted from the study SQL and LIMS databases as comma delimited files.

## 2.9 Ethical Approval, Consent and Participant Remuneration

The study protocol was granted ethical approval by the Malawi College of Medicine Research Ethics Committee (COMREC), protocol number P.11/16/2063 and the Liverpool School of Tropical Medicine (LSTM) Research Ethics Committee, protocol number 16-062. LSTM acted as the study sponsor. All study team member were trained in NIHR Good Clinical Practice (GCP).

All participants in the study provided informed, written consent, as follows: if a patient lacked capacity to provide informed consent themselves then informed consent was sought from the patient's representative, usually their guardian. Patients with no representative and lacking capacity to provide informed consent were not recruited. Initially, eligible patients (or their representative) were approached by a member of the study team and the study was explained to them including study procedures, risks and benefits, financial and confidentiality considerations and how to obtain more information. Written patient information leaflets were provided in English and Chichewa. If the patient/representative was willing to enter the study then they were asked to sign and date two copies of the consent form, and provided with a copy of the form to keep. If either the patient or representative was unable to read then the consent form was read to them by the study team, and witnessed by an additional staff member who was not part of the study team. If the patient (or their representative) agreed to enter the study, then the witness signed and dated the form. Any patient who was enrolled to the study after consent from a representative and who then regained the capacity for informed consent was subsequently approached independently for informed consent. A patient could withdraw at any time without giving a reason.

Hospitalised patients were not financially compensated for their time and/or transport costs, but all other participants were, following standard MLW guidelines: 500MWK (approximately USD\$0.7 at February 2019 exchange rates) was provided to participants if they were visited at home and 2000MWK (approximately USD\$2.8 at February 2019

exchange rates) or their reported transport costs (whichever was higher) for participants who attended QECH for study purposes.



## Chapter 3

# A clinical and microbiological description of sepsis in Blantyre, Malawi

### 3.1 Chapter overview

In this chapter, I present a clinical and microbiological description of sepsis in Blantyre, Malawi. As expected, participants are young relative to high income settings and there is a high prevalence of HIV-infection. 28-day case-fatality is 18% (95% CI 13-26%) but continues to rise throughout the study period to 31% (95% CI 25-38%) by 180 days and is higher in the HIV-infected in a time-to event analysis, seemingly driven by late (> 2 week) deaths. Participants had been unwell for some time (median 7 days) prior to presentation. Microbiological testing identified an aetiological agent for 51% of the cohort: *Mycobacterium tuberculosis* is the commonest cause identified, in 34% of all participants, usually manifesting as disseminated mycobacterial disease diagnosed by urinary lipoarabinomannan testing. Bloodstream infection was next most prevalent (11%), followed by malaria caused by *Plasmodium falciparum* (9%). I use logistic regression to identify determinants of 28-day mortality. Time to antibacterial therapy and volume of intravenous fluid administered show no significant association with mortality; however, receipt of antituberculous showed an independent association with survival to 28 days (aOR 0.17 for death [95% CI 0.03-0.74]). The logistic regression models used to provide these estimates of effect size had some problems, which could result in biased estimates: some predictors perfectly predicted outcome (i.e. separation occurred) and so were excluded from the model, and the number of predictors included meant that the model is likely overparametrised.

## 3.2 Introduction and chapter aims

The aetiology of sepsis in sSA is poorly described (reviewed in Chapter 1), and the appropriateness of the blind introduction of sepsis treatment principles and protocols from high-income settings to sSA is coming under increasing scrutiny. A focus on reducing door-to-antimicrobial time and rapid administration of intravenous fluid is likely effective at reducing sepsis mortality in high-income settings[6,7], but data on the utility of this strategy in sSA are limited, and it is possible that other patient or treatment factors may be stronger determinants of outcome in sSA. The aims of this chapter are twofold:

1. To describe the demographics, mode of presentation, aetiology, management, and outcomes of sepsis in Blantyre with to 180 days in terms of mortality and morbidity as health related quality of life.
2. To identify independent associations of mortality - particularly modifiable treatment-related factors - with a view to informing sepsis treatment protocols for sSA.

## 3.3 Methods

The clinical and laboratory methods of the clinical study are described in Chapter 2, Methods; a further overview of chapter aims and description of the statistical analysis used is given here.

For the first aim - description of sepsis in Blantyre - patient demographics, health seeking behaviour, symptoms and admission physiology are described as medians and interquartile ranges for continuous variables or proportions for categorical variables. Difference in medians or difference in proportions with bootstrapped or binomial 95% confidence intervals, respectively, were used to assess for difference between groups. For bootstrapped differences in medians 1000 replicates were used and confidence intervals constructed using the percentile bootstrap. Aetiology is presented as simple proportions with exact binomial 95% confidence intervals, stratified by HIV status and co-infections visualised as Venn diagrams and UpSet plots using *eulerr*[292] and *UpSetR*[293] packages, respectively, in R. Mortality is presented as simple proportions at 28, 90 and 180 days with exact binomial confidence intervals, and Kaplan-Meier estimates of the survival function generated using the *survival* package in R[294]. Both of these estimates are presented aggregated and stratified by HIV status, with difference in proportions with 95% binomial confidence intervals used to test the hypothesis that point mortality estimates differ between HIV-infected and uninfected participants, and log-rank test used to test the hypothesis that HIV-infected and uninfected survival functions differ.

Morbidity was assessed as health related quality of life (HRQoL) using the EQ-5D-3L

questionnaire, which assesses HRQoL across five domains (anxiety, pain, self care, usual activities, walking) with participants describing their problems across the domains on a 3 point ordinal scale: no problems, moderate or extreme. The level of disability is selected by the participant by choosing from the statements: “I have no problems”, “I have some problems” or “I have extreme problems” for each domain. So-called tariff sets are used to convert these scores to an overall utility score, which compares the health state compared to perfect health: a utility score of one represents perfect health and zero represents death, but scores of below zero (states worse than death) are possible. Tariff sets are country specific, and no Malawian tariff sets are available, so a Zimbabwean tariff set was used[295]. The *eq5d* package in R was used to convert health states into utility scores. HRQoL was measured at baseline, and the 1,4,12 and 24 week visits, and is presented as proportion of participants reporting at least moderate impairment in each domain (with exact binomial confidence intervals) at each time point, as well as boxplots of utility score. Utility scores of participants with sepsis were compared to community controls using difference in means and 95% confidence intervals between the groups.

Univariable associations of 28-day mortality were assessed by 1) calculating differences in medians or proportions between participants who died and survived, with 95% confidence intervals constructed from the Chi-square distribution for proportions or percentile bootstrap as above and 2) from univariable logistic regression models considering each variable alone as a predictor. Variables were selected for these comparisons that *a priori* may be expected to be associated with mortality: host variable (age, sex, HIV status, CD4 cell count, ART status, haemoglobin), severity variables (admission vital signs, inability to stand on admission, white cell count (WCC), platelet count, serum creatinine, urea, and bicarbonate, capillary lactate), diagnosis, or treatment variables (receipt of any antibacterial, any antifungal, any antituberculous drug or any antimalarial all as binary variables, time to receipt of each as a continuous variable, and volume of intravenous fluid received in the first 6 hours of admission as a continuous variable). To address confounding, multivariable logistic regression models were constructed, using all these variables except CD4 cell count, ART status and time-to-antibacterials (as these variables were not available for all participants). Otherwise the full set of *a priori* variables were used as predictors and 28-day mortality as the outcome, and fitted using the *glm* command in R v3.6.0. All statistical analysis were carried out in R v3.6.0.

## 3.4 Results

### 3.4.1 Study population

225 participants were recruited in 20 months between 19th February 2017 and 2nd October 2018 (Figure 3.1). Participants were recruited soon after arrival in hospital (median 1.5 hours [IQR 0.8-2.6] after first attendance). In total, 4 participants (2%) were lost to follow up over the 180-day study period; 5 participants (2%) withdrew; and 7 participants (3%) transferred out of the study area before 180 days. Four of the five participants who withdrew gave a reason for their wish to withdraw, all that they no longer wished the inconvenience of being involved in the study. 15/225 (7%) participants had their final study visit before 180 days, and so were not included in the 180-day outcome analysis.

### 3.4.2 Baseline characteristics

Table 3.1 shows the baseline characteristics of the recruited participants. Subjects tended to be young (median [IQR] age 36 [28-44]). The majority were HIV-infected (67% [143/213] of those with known HIV status), of whom 117 (82%) were on ART, almost exclusively the Malawian first-line regimen of efavirenz, lamivudine and tenofovir. The majority (88/117 [75%]) had been taking ART for more than three months; the distribution of reported time on ART is shown in Figure 3.2.

Almost all (221/225 [98%] of participants) experienced subjective fever (Figure 3.3). Participants had been unwell for some time, a median (IQR) of 7 (3-14) days; 32/225 (14%) of participants had been unwell for more than 4 weeks. 18/225 (8%) of participants had been admitted to hospital within the last 4 weeks. Over half (123/225 [55%]) of participants had sought care for their current illness (Table 3.2), most commonly at a government health centre (101/123 [82%]), a median (IQR) of 2 (1-6) days previously.

Prehospital antimicrobial use was common: 60/225 (27%) of all participants had received an antimicrobial for their current illness: 7/60 (12%) of all prehospital antimicrobials were antimalarials, the remainder antibacterial, most commonly co-trimoxazole or ciprofloxacin. Prehospital intravenous or intramuscular antimicrobials were administered in 16/60 (27%) participants receiving antimicrobials: ceftriaxone (n=6), benzylpenicillin (n=4), gentamicin (n=3) and artesunate (n=3).



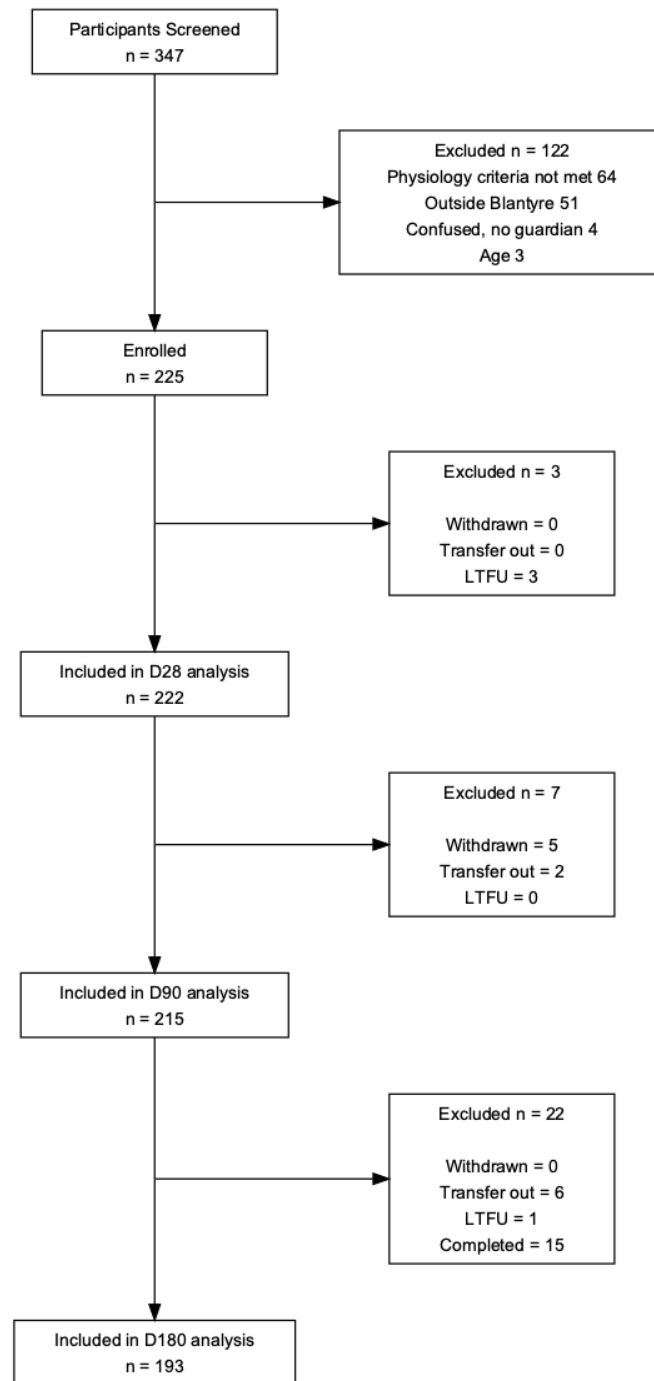


Figure 3.1: Study recruitment and follow up.

Table 3.1: Participant characteristics

Variable	Value
<b>Demographics</b>	
Age (years)	36 (28-44)
Male sex	114/225 (51% [44-57%])
<b>HIV/TB status</b>	
HIV Reactive	143/225 (64% [57-70%])
HIV Non Reactive	70/225 (31% [25-38%])
HIV Unknown	12/225 (5% [3-9%])
Ever treated for TB	37/225 (16% [12-22%])
Of those, current TB treatment	10/37 (27% [14-44%])
<b>ART status*</b>	
Current ART	117/143 (82% [75-88%])
Months on ART	29 (4-73)
ART regimen: EFV/3TC/TDF	110/117 (94% [88-98%])
ART regimen: other	7/117 (6% [2-12%])
Current CPT <sup>†</sup>	98/141 (70% [61-77%])
<b>Tobacco/alcohol use</b>	
Never tobacco	196/225 (87% [82-91%])
Ex tobacco	17/225 (8% [4-12%])
Current tobacco	12/225 (5% [3-9%])
Current alcohol	51/225 (23% [17-29%])
<b>Employment</b>	
Unemployed	82/225 (36% [30-43%])
Currently employed	65/225 (29% [23-35%])
Self-employed	56/225 (25% [19-31%])
Student	21/225 (9% [6-14%])
Retired	1/225 (0% [0-2%])
<b>Animals at home?</b>	
Any animal	71/225 (32% [26-38%])
Poultry	46/71 (65% [53-76%])
Dogs	18/71 (25% [16-37%])
Goats	12/71 (17% [9-28%])
Dogs	18/71 (25% [16-37%])
Other	11/71 (15% [8-26%])

*Note:*

ART = Antiretroviral therapy, CPT = Co-trimoxazole preventative therapy, EFV: Efavirenz, 3TC: Lamivudine, TDF: Tenofovir. Numeric values are median (IQR) or proportions with exact binomial 95% confidence intervals unless otherwise stated.

\* ART status includes HIV reactive only as denominator

<sup>†</sup> Missing CPT data for two participants.

### 3.4.3 Admission physiology and laboratory investigations

Admission vital signs and laboratory investigations are shown in Table 3.3. Shock, defined as systolic blood pressure below 90mmHg, was more common than hypoxia, defined as oxygen saturation below 90% (34% [80/225] vs 13% [28/224]). Reduced conscious level,

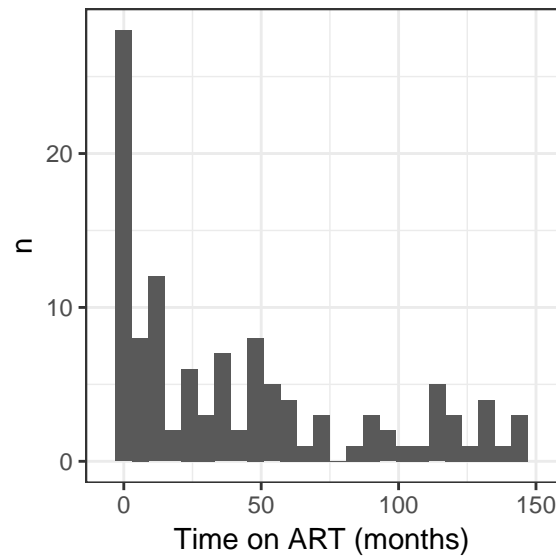


Figure 3.2: Distribution of reported time of ART (months). Histogram bins are 6 months

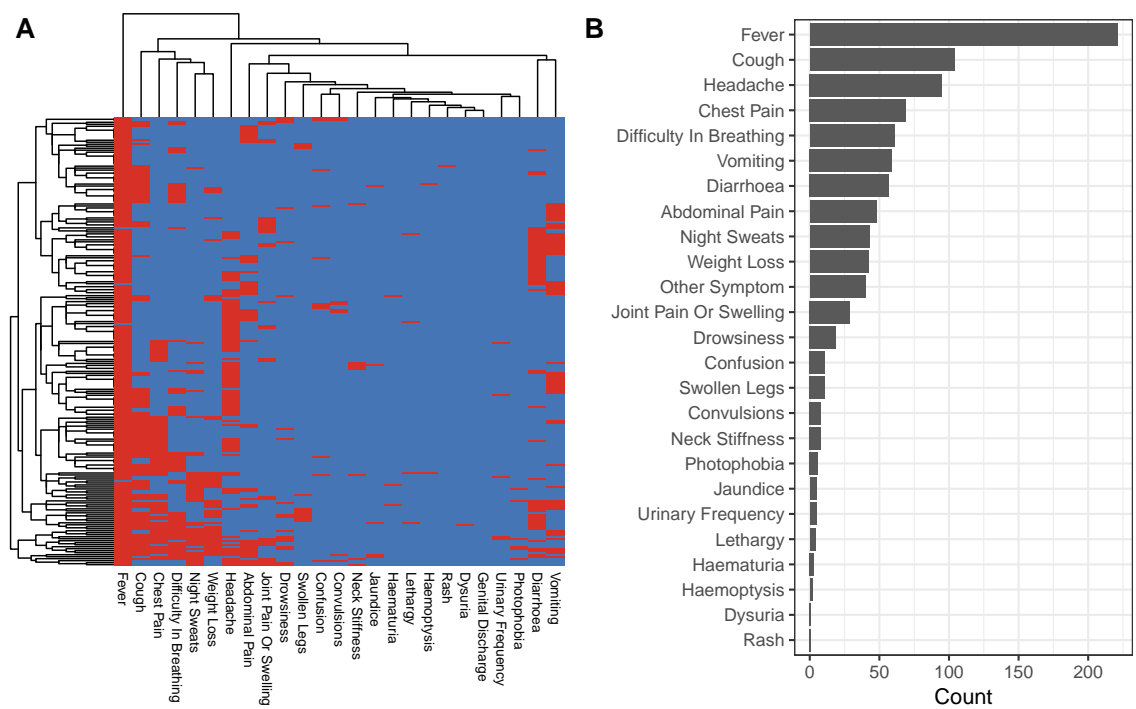


Figure 3.3: Symptoms of recruited participants. A: Row and column clustered heatmap of participant symptoms. Each row represents a patient. Red = presence, blue = absence. B: Frequency of occurrence of symptoms

Table 3.2: Prehospital healthcare seeking and antimicrobial exposure

Variable	Value
<b>Pre-hospital healthcare seeking</b>	
Sought care prior to attendance at hospital	123/225 (55% [48-61%])
At health centre	101/123 (82% [74-88%])
At hospital	16/123 (13% [8-20%])
At private doctor	8/123 (7% [3-12%])
Somewhere else	1/123 (1% [0-4%])
Days prior to today that participant sought care	2 (1-6)
<b>Prehospital antimicrobial exposure</b>	
Received any antimicrobial prior to attendance at hospital	60/225 (27% [21-33%])
Co-trimoxazole	12/60 (20% [11-32%])
Ciprofloxacin	10/60 (17% [8-29%])
Amoxicillin	9/60 (15% [7-27%])
Ceftriaxone	6/60 (10% [4-21%])
Metronidazole	5/60 (8% [3-18%])
Benzylpenicillin	4/60 (7% [2-16%])
Artesunate	3/60 (5% [1-14%])
Gentamicin	3/60 (5% [1-14%])
Erythromycin	2/60 (3% [0-12%])
LA	2/60 (3% [0-12%])
SP	2/60 (3% [0-12%])
Azithromycin	1/60 (2% [0-9%])
Flucloxacillin	1/60 (2% [0-9%])
Days prior to today that antimicrobials started	2 (1-5)
<b>Method of transport to hospital</b>	
Minibus	78/225 (35% [28-41%])
Taxi	65/225 (29% [23-35%])
Private car/truck	42/225 (19% [14-24%])
Ambulance	37/225 (16% [12-22%])
Other	2/225 (1% [0-3%])
Walk	1/225 (0% [0-2%])
Cost (MWK) of transport to hospital	1000 (275-3000)

*Note:*

LA = Lumefantrine-artemether, SP = Sulfamethoxazole-pyrimethamine, MWK = Malawian Kwacha. Numeric values are median (IQR) or proportions with exact binomial 95% confidence intervals unless otherwise stated.

defined as Glasgow coma scale (GCS) below 15 was rarer (9% [21/225]); 28% (63/225) of participants were unable to stand on admission.

Despite high ART coverage (117/143 [82%]) among HIV-infected participants for a median of 29 months, the median (IQR) CD4 count was low at 98 (31-236) cells  $\mu\text{L}^{-1}$ . 108/141 (70%) of participants had a CD4 count below 200 cells  $\mu\text{L}^{-1}$ . CD4 count was similar in participants who had started ART more than 6 months ago as compared to less than three months ago (median [IQR] 99 [27-260] vs 93 [39-137] cells  $\mu\text{L}^{-1}$  respectively). 42/83 (51%) of participants who had been taking ART for more than 6 months had a CD4 count of less than 100 cells  $\mu\text{L}^{-1}$ , which fulfils the WHO definition of immunological failure of HIV treatment.

Table 3.3: Admission physiology, haematology and biochemistry

Variable	Value
<b>Admission physiology</b>	
Temperature ( $^{\circ}\text{C}$ )	38.5 (37.9-39.0)
Heart rate ( $\text{min}^{-1}$ )	121 (102-132)
Systolic BP (mmHg)	99 (85-119)
Diastolic BP (mmHg)	66 (56-76)
MAP (mmHg)	76 (65-89)
Respiratory rate ( $\text{min}^{-1}$ )	34 (32-38)
Oxygen saturation (%)	96 (94-98)
GCS 15	204/225 (91% [86-94%])
GCS 11-14	16/225 (7% [4-11%])
GCS < 11	5/225 (2% [1-5%])
Unable to stand	63/225 (28% [22-34%])
<b>Admission CD4 count</b>	
CD4 count* ( $\mu\text{L}^{-1}$ )	98 (31-236)
<b>Admission haematology</b>	
Haemoglobin ( $\text{x}10^9 \text{ g dL}^{-1}$ )	10.8 (8.2-13.2)
White cell count ( $\text{x}10^9 \text{ L}^{-1}$ )	6.5 (4.4-11.4)
Neutrophil count ( $\text{x}10^9 \text{ L}^{-1}$ )	4.0 (2.1-7.5)
Platelet count count ( $\text{x}10^9 \text{ L}^{-1}$ )	218 (146-297)
<b>Admission biochemistry</b>	
Sodium ( $\text{mmol L}^{-1}$ )	134 (130-137)
Potassium ( $\text{mmol L}^{-1}$ )	4.0 (3.6-4.4)
Bicarbonate ( $\text{mmol L}^{-1}$ )	19 (17-22)
Chloride ( $\text{mmol L}^{-1}$ )	101 (97-104)
Urea ( $\text{mmol L}^{-1}$ )	4.8 (3.5-8.0)
Creatinine ( $\text{mmol L}^{-1}$ )	76 (59-103)
Lactate ( $\text{mmol L}^{-1}$ )	3.4 (2.3-5.2)

*Note:*

GCS = Glasgow coma scale, BP = Blood pressure, MAP = Mean arterial blood pressure. Numeric values are median (IQR) or proportions with exact binomial 95% confidence intervals unless otherwise stated.

\* CD4 count includes only HIV-infected participants; 2 values were missing.

### 3.4.4 Aetiology

In total, 51% (114/225) of the 225 participants had at least one infectious agent identified (Table 3.4), most commonly disseminated mycobacterial infection (76/225 [34%]) followed by culture confirmed bacterial or fungal bloodstream infection (24/225 [11%]) and *Plasmodium falciparum* malaria (21/225 [9%]). Table 3.5 shows the availability of test and proportion of positive tests across the cohort, stratified by HIV status. 2/225 patients (1%) had a missing aerobic blood culture; the remaining 223 patients had a total of 259 blood cultures performed. 15/259 (6%) blood cultures grew at least one contaminant, but 26 blood cultures from 24 patients were positive for a total of 28 pathogenic bacteria (Figure 3.6): *Salmonella enterica* Seroovar Typhi was the most commonly isolated bacterial pathogen. Of note 8/9 participants from whom *S. Typhi* was isolated were HIV noninfected, with one untested. Of the 18 Gram-negative bacteria isolated, 3/18 (17%) were cefpodoxime resistant on AST via disc diffusion testing, and likely ESBL producers: one *K. pneumoniae* and one *E. coli* (both from the same blood culture and same patient) and one *Acinetobacter baumannii*. Both *Staphylococcus aureus* isolates were oxacillin sensitive. The one *Streptococcus pneumoniae* cultured was penicillin intermediate on AST. There was no significant difference in proportion with positive blood culture in those who self reported receipt of prehospital antibacterials versus those who did not (7/52 [13%] BSI in those reporting prehospital antibacterials vs 17/154 [11%] not, difference 2% [95% CI -9-14%]).

Lumbar puncture and CSF culture was carried out in 44 participants: 5/44 (11%) of samples grew a containment and no pathogenic bacteria were recovered from any sample. 4/44 (9%) had a detectable cryptococcal antigen (CRAG) in CSF. Malaria testing was missing for 6/225 (3%) of participants, but of the remainder, a positive malaria test was more likely in the HIV-uninfected (12/69 [17%] vs 6/138 [4%], difference 13% [95% CI 2-24%]). Positive aerobic blood culture showed no statistically significant association with HIV, nor did positive CSF testing, though in the latter case numbers were small and all positive tests (all positive CRAG) were in fact in the HIV-infected (Table 3.5).

Testing for TB, with the exception of sputum Xpert testing, was restricted to HIV-infected participants. Sputum Xpert was carried out in 44/225 (20%) of participants, and was more commonly carried out in the HIV-infected: 35/143 [24%] of HIV-infected participants had sputum testing performed vs 8/70 (11%) of HIV uninfected (difference 13% [95% CI 2-24%]). 53 sputum samples were sent in total from the 44 patients, and 8/44 (18%) diagnoses of TB made, all except one in HIV-infected participants. One sample identified a rifampicin resistance gene, but the participant died before the result was available and so was not started on MDR-TB treatment; the remainder of infections were rifampicin-sensitive.

155 participants were eligible for urinary lipoarabinomannan (uLAM) and mycobacterial blood culture testing, being either HIV-infected (n=143) or of unknown HIV status (n=12).

Table 3.4: Number of diagnoses

Diagnosis	Proportion of participants
Tuberculosis	76/225 (34% [28-40%])
Bloodstream infection	24/225 (11% [7-15%])
Malaria	21/225 (9% [6-14%])
Meningitis	4/225 (2% [0-4%])
No diagnosis	111/225 (49% [43-56%])

Urine was available for 145/155 (94%) of those eligible, and 74/145 (51%) of samples were positive for uLAM. 150/155 (97%) of eligible participants had blood samples collected and cultured for mycobacteria. 12/150 (8%) grew contaminants and are excluded from the denominators in Table 3.5; of the remainder 8/138 (6%) grew mycobacteria, all *M. tuberculosis*.

Figures 3.4 and 3.5 show the overlap of positive tests from the different diagnostic modalities. Of the 114 patients with at least one positive diagnostic test, 90/114 (79%) had only one positive diagnostic test. The exceptions to this were mycobacterial blood culture and sputum Xpert: patients who had TB diagnosed by these tests tended to also have a positive uLAM. 2/4 (50%) of patients with positive CSF testing (all of whom had detectable CRAG) had also grew *Cryptococcus neoformans* in aerobic blood culture. 111/225 (49%) of patients remained with no diagnosis.

Table 3.5: Diagnostic tests performed and results, stratified by HIV status.

type	HIV status		
	Reactive	Non reactive	Unknown
Number of participants	143	70	12
<b>TB diagnostics</b>			
Urinary LAM	70/136 (51% [43-60%])	-	4/9 (44% [14-79%])
Sputum Xpert	7/35 (20% [8-37%])	1/8 (12% [0-53%])	0/1 (0% [0-98%])
TB blood culture	7/128 (5% [2-11%])	-	1/10 (10% [0-45%])
<b>Other diagnostics</b>			
Aerobic blood culture	13/141 (9% [5-15%])	9/70 (13% [6-23%])	2/12 (17% [2-48%])
CSF culture or CRAG	4/31 (13% [4-30%])	0/12 (0% [0-26%])	0/1 (0% [0-98%])
Malaria RDT	6/138 (4% [2-9%])	12/69 (17% [9-28%])	3/12 (25% [5-57%])
<b>Total with diagnosis</b>	<b>86/143 (60% [52-68%])</b>	<b>21/70 (30% [20-42%])</b>	<b>7/12 (58% [28-85%])</b>

*Note:*

LAM = Lipoarabinomannan, CSF = Cerebrospinal fluid, CRAG = Cryptococcal antigen, RDT = Rapid diagnostic test. Urinary LAM and TB blood culture were not carried out in HIV negative participants.



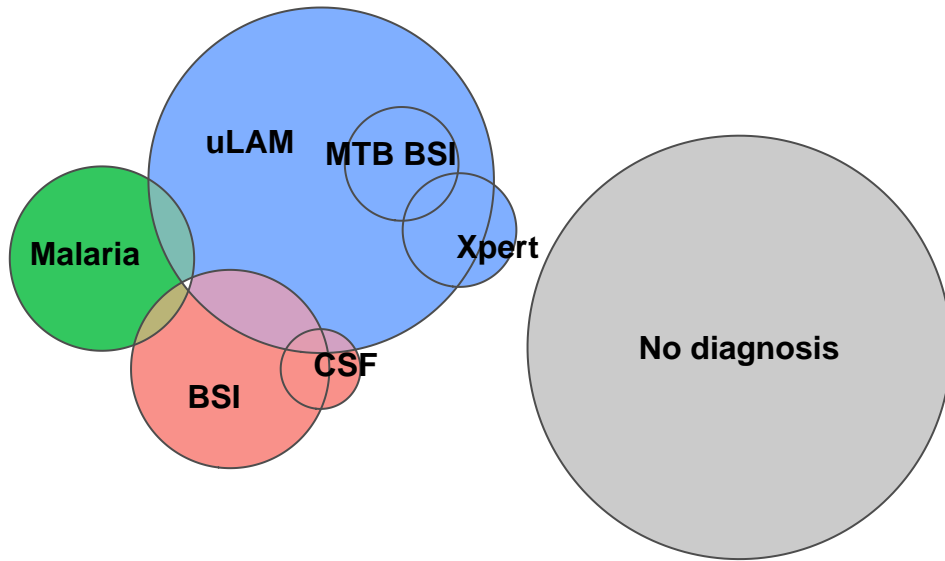


Figure 3.4: Venn diagram showing overlap of positive diagnostic tests; culture of blood and CSF shown in red, malaria in green and TB diagnostics in blue. The CSF variable includes either a positive culture for a pathogenic bacteria or positive cryptococcal antigen, BSI a positive aerobic culture of pathogenic bacteria from blood and MTB BSI culture of *M. tuberculosis* from blood. BSI: Bloodstream infection, CSF: Cerebrospinal fluid, mRDT: Malaria rapid diagnostic test, MTB BSI: *Mycobacterium tuberculosis* bloodstream infection, uLAM: urinary lipoarabinomannan.

### 3.4.5 Treatment

At least one antimicrobial drug was received by 214/225 (95%) of the cohort during their admission (Table 3.6), most commonly an antibacterial (207/225 [92%]), but also a significant minority received antituberculous chemotherapy (63/225 [28%]). Of those receiving antituberculous therapy, 10/63 (16%) were taking the medication prior to admission, and the remainder were initiated on therapy during admission. The first antibacterial agent administered was most often ceftriaxone, (181/207 [87%]) but ciprofloxacin (9%), amoxicillin (3%) and metronidazole (1%) were also used. Median door to antimicrobial time was 5.3 hours (IQR 3.7-10.8) for antibacterials and 4.5 hours (IQR 3.1-21.7) for antimalarials but longer for antifungals at 47.7 hours (IQR 27.9-73.9) and longer still for antitubercular therapy at 120.9 hours (IQR 63.7-171.0). Cumulative incidence curves for administration of the different antimicrobial classes are shown in Figure 3.7A-D.

Of all participants, 85% (192/225) received any intravenous fluid in the first 6 hours of enrolment to the study; of these, most received 0.9% saline (160/192 [83%] of those receiving fluid) but 5% dextrose (91/192 [57%]) were also used; Ringer's lactate (6/192 [6%]) and blood (2/192 [1%]) were rarely administered. Of the 192 patients who were administered any fluid, a median of 1.5L (IQR 1.0-2.0L) was administered over the 6hr study period; fluid administration as a function of time is shown in Figure 3.7E.

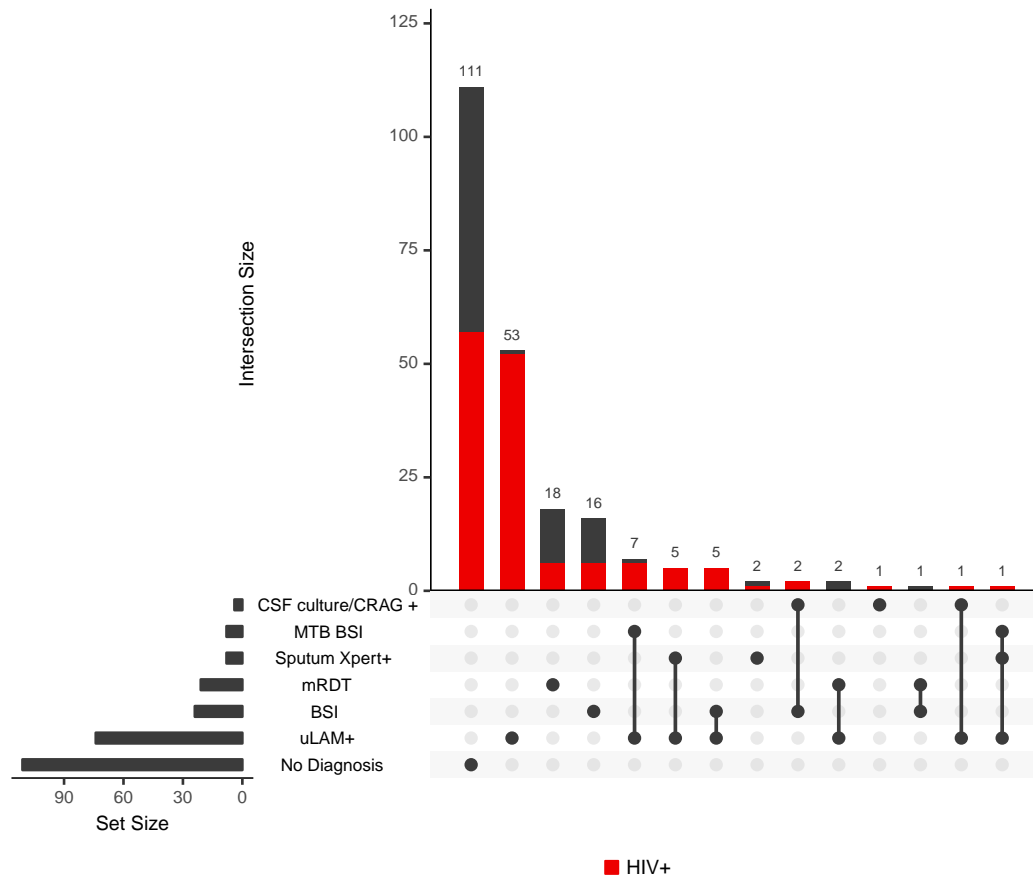


Figure 3.5: UpSet plot of overlap of positive diagnostic tests, showing that for the majority of participants, one test alone is positive. Red colour indicates HIV-infected; black is a composite of HIV-negative and unknown. The CSF variable includes either a positive culture for a pathogenic bacteria or positive cryptococcal antigen, BSI a positive aerobic culture of pathogenic bacteria from blood and MTB BSI culture of *M. tuberculosis* from blood. BSI: Bloodstream infection, CSF: Cerebrospinal fluid, CRAG: Cryptococcal antigen, mRDT: Malaria rapid diagnostic test, MTB BSI: *Mycobacterium tuberculosis* bloodstream infection, uLAM: urinary lipoarabinomannan.

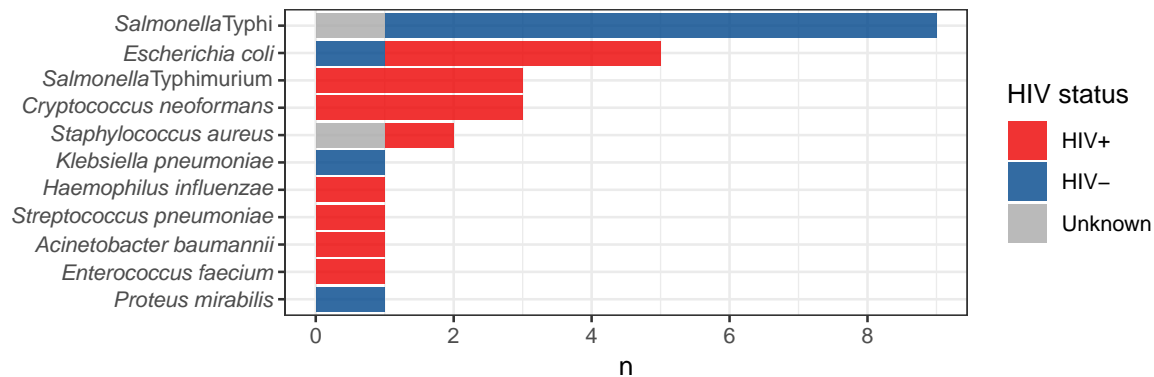


Figure 3.6: Pathogenic isolates recovered from aerobic blood culture. 26 blood cultures in 24 participants were positive for 28 pathogens in total.

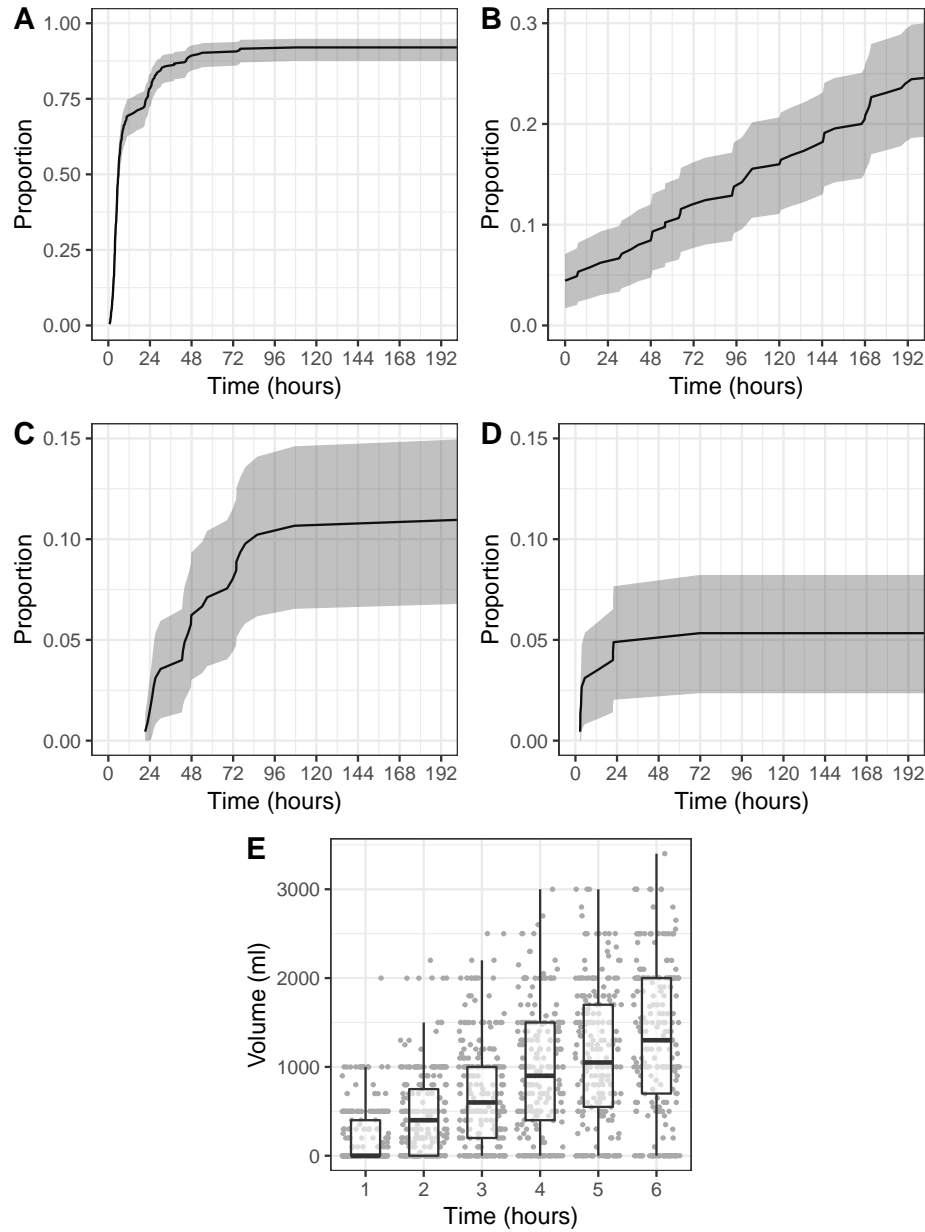


Figure 3.7: Timing of antimicrobial and fluid administration. A-D: Cumulative incidence of administration of antibacterial (A), antitubercular (B), antifungal (C) and antimalarial (D) agents as a function of time since arrival at hospital in hours. E: Total volume of administered intravenous fluid as a function of time since enrollment to study in hours. Boxplots show median, quartiles box and 1.5 times interquartile range as whiskers. Points are jittered around the hour at which they were measured to show distribution.

Table 3.6: Door-to-antimicrobial times.

Antimicrobial class	No. participants	Median [IQR] time (hours)
Antibacterial	207/225 (92% [88-95%])	5.3 (3.7-10.8)
Antitubercular	63/225 (28% [22-34%])*	120.9 (63.7-171.0)
Antifungal	26/225 (12% [8-16%])	47.7 (27.9-73.9)
Antimalarial	12/225 (5% [3-9%])	4.5 (3.1-21.7)

\* 10/63 participants who received antitubercular agents during admission were taking them prior to admission; they are excluded from the calculation of median door-to-antimicrobial time for this class.

### 3.4.6 Outcome

Median hospital stay was 4 (IQR 1-9) days. Case fatality rate (CFR) was 18% (95% CI 13-23%) at 28 days, 24% (95% CI 18-30%) at 90 days and 31% (95% CI 25-38) at 180 days, and higher in HIV-infected participants at each time point (Table 3.7), though not statistically significant: HIV-infected vs noninfected 19% vs 13% at 28 days (difference 6% [95% CI -6-13%]), 27% vs 17% at 90 days (difference 10% [95% CI -4-22%]) and 36% vs 21% at 180 days (difference 15% [95% CI 0-30%]). The primary planned Kaplan-meier estimation of the survival function (Figure 3.8) showed a precipitous decline in survivorship to around day 30 and CFR declined thereafter, to the end of the study period. Stratifying the analysis by HIV status revealed that early deaths (within the first 1-2 weeks) occur at similar rates in the two groups before the curves diverge; log-rank test suggested a significant difference in survival function between the two groups ( $p = 0.03$ ). In view of the fact that a number of the HIV-infected participants were likely to be failing ART, only two were recorded as switching to second line therapy during the study period, though details of any interventions (e.g. ART adherence counselling) or HIV viral load testing carried out as part of routine care were not captured.

Health related quality of life measures, as assessed by EQ-5D-3L, are shown in Figure 3.9 for participants with sepsis and the community cohort as a comparator. Acutely, participants with sepsis reported were significantly disabled, reporting at least moderate impairment across all domains in the majority of cases, and over 90% of participants reporting at least moderate impairment in activities of daily living and experiencing at least moderate pain or discomfort. However, recovery following treatment in survivors was rapid. The mean EQ-5Q utility score (a measure of the weight compared to a health state compared to 1, perfect health) of healthy community controls was 0.910 (SD 0.102) at enrolment, significantly higher than participants with sepsis at enrolment (utility score 0.496 (SD 0.251), difference in means versus community enrolment 0.414 [95% CI 0.364-0.463]), but comparable to participants with sepsis at their 12 week assessment (0.913 (SD 0.147), difference in means versus community enrolment scores 0.003 [95% CI -0.039-0.045]).

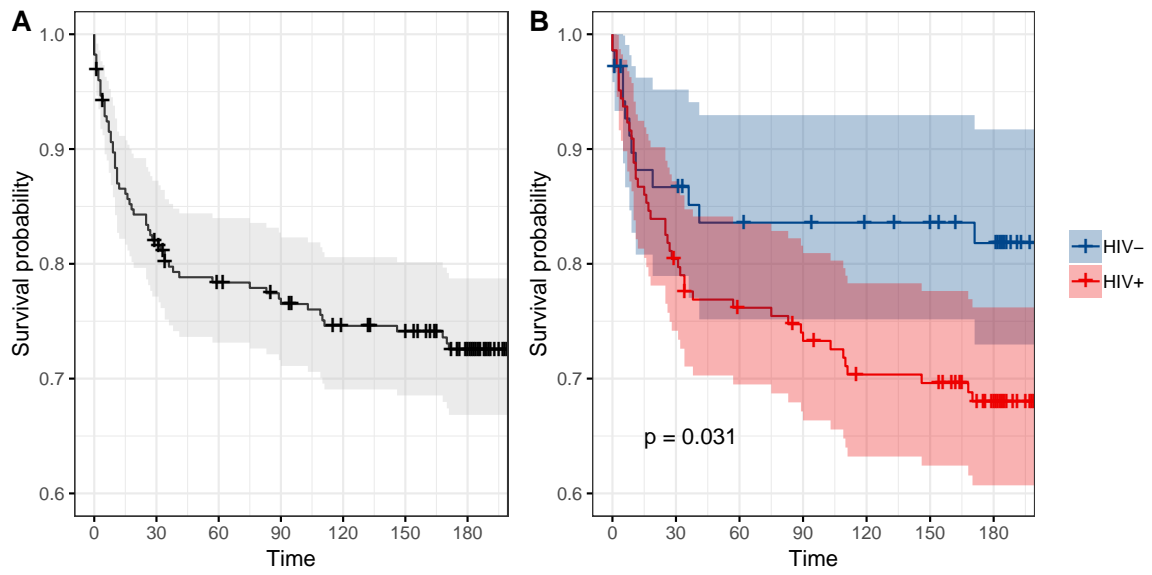


Figure 3.8: Kaplan-Meier survival curves of all included participants (A) and stratified by HIV status (B). Crosses indicate censoring. p value from log-rank test comparing survival of HIV-infected to HIV-noninfected participants shown ( $p = 0.03$ ).

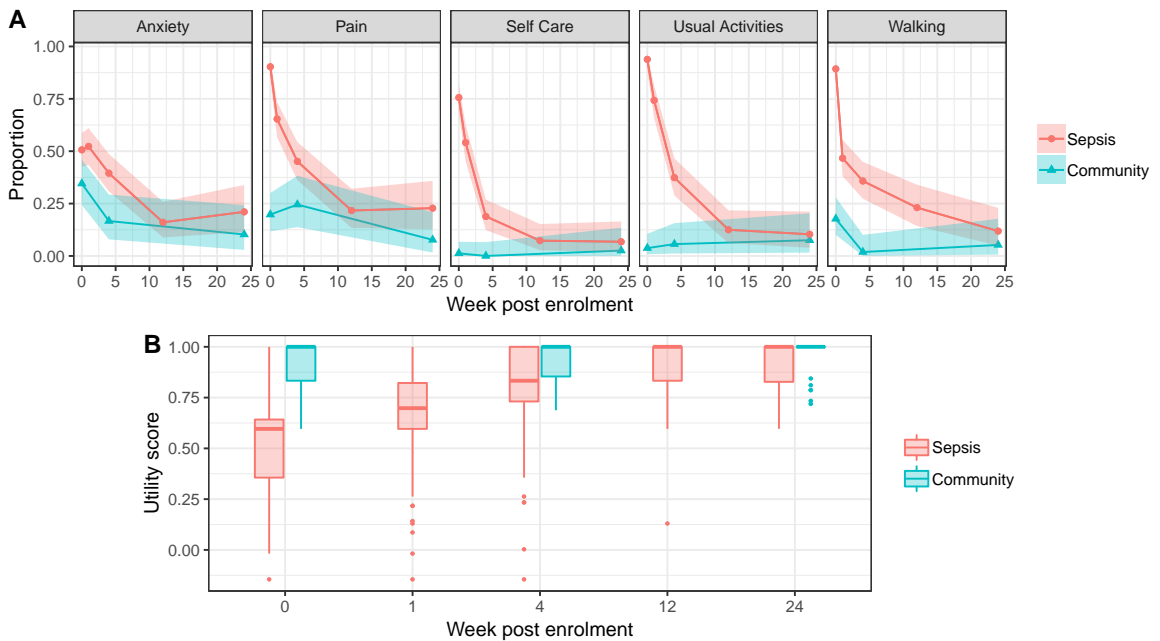


Figure 3.9: Health-related quality of life following sepsis admission, compared to community controls, showing a return to usual quality of life by 12 weeks following admission. A: proportion of participants across each of the five domains of the EQ-5D questionnaire who report at least moderate impairment. B: EQ-5D utility score derived using the Zimbabwean tariff set. The utility score is interpreted as the weight attached to a health state compared to perfect health, which is assigned a value of 1. By 12 weeks there is no statistically significant difference between sepsis and baseline community participant utility scores (difference in means 0.003 [95% CI -0.039-0.045]).

Table 3.7: Day 28, 90 and 180 mortality stratified by HIV status

	HIV+		HIV-		HIV Unknown		Total	
	n	Mortality	n	Mortality	n	Mortality	n	Mortality
Day 28	143	19% (13-26%)	67	13% (6-24%)	12	25% (5-57%)	<b>222</b>	<b>18% (13-23%)</b>
Day 90	139	27% (19-35%)	64	17% (9-29%)	12	25% (5-57%)	<b>215</b>	<b>24% (18-30%)</b>
Day 180	125	36% (28-45%)	58	21% (11-33%)	11	27% (6-61%)	<b>194</b>	<b>31% (25-38%)</b>

*Note:*

n in this table indicates the number of patients with known vital status, contributing data at the given time point (i.e. not lost to follow up, withdrawn, or transferred out).

### 3.4.7 Determinants of mortality

Bivariable associations of mortality are shown in Table 3.8 with variables grouped into putative host, severity, diagnosis and treatment variables. Variables associated with immunosuppression - CD4 count and haemoglobin - were associated with death in unadjusted analysis, as were well recognised markers of disease severity: shock, hypoxia, reduced conscious level, hyperlactataemia, and inability to ambulate, as were reduced venous bicarbonate and increased venous urea. A diagnosis of malaria was strongly associated with survival (0/21 died), as was receipt of antimalarials. Conversely, a diagnosis of meningitis was associated with mortality (Table 3.8) though numbers were small ( $n = 4$ ) and confidence intervals of differences between groups crossed zero. Receipt of antibacterials, or antifungals showed no association with mortality, though almost all participants received antibacterials and only a minority antifungals, so moderate effect sizes would be unlikely to be detected. However, receipt of antituberculous chemotherapy therapy was associated with survival: 8% (4/53) of participants receiving TB therapy died compared to 21% (35/169) who did not receive it (difference 13% [95% CI 3-30%]).

All of these associations are likely to be affected by confounding, so I constructed logistic regression models to attempt to produce unbiased effect estimates (Table 3.9). There were some difficulties with these models. Separation occurred when malaria and meningitis diagnoses and receipt of antimalarials were included - they perfectly predicted outcome, so parameter estimates become unstable - and so they were excluded from the model. Receipt of TB therapy remained strongly associated with survival (aOR 0.17 [95% CI 0.03-0.74]) as did higher haemoglobin (aOR 0.69 [95% CI 0.52-0.86] per  $1\text{g dL}^{-1}$  increase) and higher oxygen saturation (aOR 0.67 [95% CI 0.47-0.96] per 5% increase). Perhaps surprisingly, an *increased* respiratory rate was associated with survival (aOR 0.35 (0.14-0.81) per 10 breaths  $\text{min}^{-1}$  increase; less surprisingly, inability to stand on admission strongly predicted mortality (aOR 14.55 [95% CI 3.81-69.78]).

Table 3.8: Bivariate associations with death by 28 days

Variable	Died	Survived	Difference
Host Variables			
Age (years)	36.4 (31.5-46.0)	35.9 (27.4-42.9)	-0.5 (-7.9 to 3.3)
Male sex	19/39 (49%)	93/183 (51%)	2% (-17 to 21%)
HIV Infected*	27/36 (75%)	116/174 (67%)	-8% (-26 to 9%)
Taking ART <sup>†</sup>	21/27 (78%)	96/116 (83%)	5% (-14 to 24%)
<b>CD4 count<sup>†</sup> (<math>\mu\text{L}^{-1}</math>)</b>	<b>28.5 (9.5-124.5)</b>	<b>103.0 (43.5-251.0)</b>	<b>74.5 (1.5 to 108.5)</b>
<b>Haemoglobin (<math>\times 10^9</math> g dL<sup>-1</sup>)</b>	<b>9.1 (6.0-10.4)</b>	<b>11.0 (8.6-13.4)</b>	<b>1.9 (0.8 to 4.0)</b>
Severity Variables			
<b>Temperature (<math>^{\circ}\text{C}</math>)</b>	<b>38.1 (37.7-38.8)</b>	<b>38.5 (38.0-39.0)</b>	<b>0.4 (0.0 to 0.7)</b>
Heart rate ( $\text{min}^{-1}$ )	123.0 (104.5-138.5)	120.0 (102.0-131.0)	-3.0 (-13.0 to 6.0)
<b>Systolic BP (mmHg)</b>	<b>89.0 (75.0-103.0)</b>	<b>99.0 (86.5-118.5)</b>	<b>10.0 (1.0 to 20.0)</b>
<b>Diastolic BP (mmHg)</b>	<b>58.0 (49.5-70.5)</b>	<b>67.0 (57.0-75.5)</b>	<b>9.0 (2.0 to 16.0)</b>
<b>Mean arterial BP (mmHg)</b>	<b>68.3 (59.7-80.7)</b>	<b>78.7 (67.0-89.2)</b>	<b>10.3 (3.7 to 17.3)</b>
Respiratory rate ( $\text{min}^{-1}$ )	34.0 (32.0-36.5)	34.0 (32.0-38.0)	0.0 (-2.0 to 2.0)
<b>Oxygen saturation (%)</b>	<b>95.0 (89.5-97.0)</b>	<b>97.0 (95.0-98.0)</b>	<b>2.0 (0.0 to 3.0)</b>
GCS	15.0 (15.0-15.0)	15.0 (15.0-15.0)	0.0 (0.0 to 0.0)
<b>Unable to stand</b>	<b>27/39 (69%)</b>	<b>36/183 (20%)</b>	<b>-50% (-67 to -32%)</b>
<b>Lactate (mmol L<sup>-1</sup>)</b>	<b>4.9 (3.0-10.6)</b>	<b>3.2 (2.1-4.5)</b>	<b>-1.7 (-6.3 to -0.1)</b>
White cell count ( $\times 10^9$ L <sup>-1</sup> )	5.9 (3.5-11.0)	6.9 (4.5-11.5)	1.0 (-1.4 to 2.8)
Platelet count ( $\times 10^9$ L <sup>-1</sup> )	181.5 (86.8-300.8)	223.0 (148.0-296.5)	41.5 (-30.0 to 72.0)
<b>Bicarbonate (mmol L<sup>-1</sup>)</b>	<b>17.0 (14.0-21.0)</b>	<b>20.0 (17.0-22.0)</b>	<b>3.0 (0.0 to 4.5)</b>
<b>Urea (mmol L<sup>-1</sup>)</b>	<b>7.8 (4.5-14.3)</b>	<b>4.5 (3.2-7.0)</b>	<b>-3.3 (-8.7 to -1.2)</b>
Creatinine (mmol L <sup>-1</sup> )	90.0 (60.0-185.0)	73.0 (59.0-96.0)	-17.0 (-48.0 to 7.0)
Diagnosis			
BSI	3/39 (8%)	20/183 (11%)	3% (-8 to 14%)
TB	15/39 (38%)	61/183 (33%)	-5% (-23 to 13%)
<b>Malaria</b>	<b>0/39 (0%)</b>	<b>21/183 (11%)</b>	<b>11% (5 to 18%)</b>
Meningitis	3/39 (8%)	1/183 (1%)	-7% (-17 to 3%)
No diagnosis	21/39 (54%)	88/183 (48%)	-6% (-25 to 13%)
Treatment Received			
Antibacterials	37/39 (95%)	167/183 (91%)	-4% (-13 to 6%)
Time to Antibacterials (hr)	4.7 (3.8-8.8)	5.3 (3.6-10.8)	0.6 (-1.0 to 1.7)
Antifungals	7/39 (18%)	19/183 (10%)	-8% (-22 to 7%)
Time to Antifungals (hr)	68.5 (45.0-72.7)	47.6 (26.6-76.4)	-20.9 (-43.2 to 14.3)

Table 3.8: Bivariate associations with death by 28 days (*continued*)

Variable	Died	Survived	Difference
<b>Antimalarials</b>	<b>0/39 (0%)</b>	<b>12/183 (7%)</b>	<b>7% (1 to 12%)</b>
Time to Antimalarials (hr)	-	4.5 (3.1-21.7)	-
<b>Antimycobacterials</b>	<b>4/39 (10%)</b>	<b>49/183 (27%)</b>	<b>17% (3 to 30%)</b>
Time to Antimycobacterials (hr)	107.3 (23.6-138.7)	99.0 (37.0-169.4)	-8.3 (-110.0 to 104.8)
IV fluid (L)	1.4 (1.0-2.0)	1.3 (0.6-2.0)	-0.1 (-0.6 to 0.2)

*Note:*

BP = Blood pressure, GCS = Glasgow coma scale. Numeric variables are presented as median (IQR) and categorical variables as proportions. Difference column shows difference in medians or difference in proportions with 95% confidence intervals. Variables shown in bold are those for which the 95% confidence intervals do not cross 0.

\* Participants with HIV status unknown not included in this row

† Includes only HIV-infected participants



Table 3.9: Unadjusted and adjusted odds ratios of death by 28 days

Variable	Univariable models	Multivariable model
	OR (95% CI)	aOR (95% CI)
Host Variables		
Age (per 5 years increase)	1.09 (0.94-1.26)	0.99 (0.76-1.27)
Male sex (vs female)	0.92 (0.46-1.84)	1.04 (0.31-3.55)
HIV Infected (vs uninfected)	1.50 (0.68-3.57)	0.31 (0.05-1.71)
HIV Unknown (vs uninfected)	2.15 (0.42-8.93)	0.22 (0.01-3.25)
Haemoglobin (per g dL <sup>-1</sup> increase)	<b>0.81 (0.72-0.90)</b>	<b>0.69 (0.52-0.86)</b>
Severity Variables		
Temperature ( per °C increase)	<b>0.61 (0.42-0.86)</b>	0.65 (0.33-1.26)
Heart rate (per 10 min <sup>-1</sup> increase)	1.02 (0.87-1.20)	1.10 (0.83-1.48)
Mean arterial BP (per 10 mmHg increase)	<b>0.77 (0.62-0.95)</b>	1.11 (0.70-1.69)
Respiratory rate (per 10 min <sup>-1</sup> increase)	0.75 (0.45-1.23)	<b>0.35 (0.14-0.81)</b>
Oxygen saturation (per 5% increase)	<b>0.72 (0.58-0.88)</b>	<b>0.67 (0.47-0.96)</b>
GCS (per 1 unit increase)	0.82 (0.65-1.01)	0.83 (0.58-1.20)
Unable to stand	<b>9.19 (4.34-20.51)</b>	<b>14.55 (3.81-69.78)</b>
Lactate (per 1 mmol L <sup>-1</sup> increase)	<b>1.27 (1.15-1.41)</b>	1.06 (0.88-1.29)
White cell count (per 1x10 <sup>9</sup> L <sup>-1</sup> increase)	0.98 (0.91-1.03)	0.97 (0.86-1.07)
Platelet count (per 100x10 <sup>9</sup> L <sup>-1</sup> increase)	0.93 (0.72-1.15)	1.11 (0.77-1.55)
Bicarbonate (per 1 mmol L <sup>-1</sup> increase)	<b>0.89 (0.82-0.96)</b>	0.93 (0.80-1.08)
Urea (per 1 mmol L <sup>-1</sup> increase)	<b>1.12 (1.06-1.19)</b>	1.05 (0.90-1.22)
Creatinine (per 10 mmol L <sup>-1</sup> increase)	<b>1.04 (1.01-1.08)</b>	1.00 (0.95-1.08)
Diagnosis		
BSI (vs no BSI)	0.68 (0.15-2.12)	0.15 (0.02-1.01)
TB (vs no TB)	1.25 (0.60-2.53)	0.94 (0.24-3.57)
Malaria (vs no malaria)	-	-
Meningitis (vs no meningitis)	<b>15.17 (1.88-311.38)</b>	-
Treatment Received		
Received any antibacterial (vs none)	1.77 (0.48-11.52)	-
Received any antifungal (vs none)	1.89 (0.69-4.70)	2.12 (0.40-11.07)
Received any antimalarial (vs none)	-	-
Received any antimycobacterial (vs none)	<b>0.31 (0.09-0.83)</b>	<b>0.17 (0.03-0.74)</b>
IV fluid received (per L increase)	1.00 (1.00-1.00)	1.00 (1.00-1.00)

*Note:*

BP = Blood pressure, GCS = Glasgow coma scale, BSI = Bloodstream infection, TB = tuberculosis. Separation occurred for those variables for which no parameter estimates are given, and they were excluded from the multivariable model. All odds ratios are for an increase in the variables shown. Variables for which the 95% confidence intervals of odds ratios do not cross 1 are shown in bold.

## 3.5 Discussion

### 3.5.1 Demographics and outcome: significant longer-term mortality

In this chapter, I have presented a clinical and microbiologic description of sepsis in adults in Blantyre, Malawi. Inkeeping with sepsis cohorts elsewhere, the participants are young, with high prevalence of HIV infection. The proportion of HIV-infected participants (67% of those with known HIV status) is comparable to a study of Sepsis-2 defined sepsis which recruited in QECH in 2008/9 (75%) but lower than sSA sepsis studies with the highest prevalence of HIV-infected participants Uganda in 2006[17] (85%) and 2009[27] (86%) and Zambia in 2012-13[9] (90%). Notably, the proportion of participants receiving HIV therapy (82%) is high compared to other sepsis studies in sSA: higher than the 08/09 Malawian study (44%), Uganda (12-24%) and Zambia (51%), which likely reflects both the success of the Malawian ART programme as well as the impressive increases in ART coverage across the continent. Despite this ostensibly high coverage, it is likely that presentation with sepsis is a manifestation of ART failure for many participants as evidenced by the low CD4 cell counts despite ART. In the significant minority of participants who recently initiated ART - 25% of those on ART started it less than three months before presentation - it seems likely that immune reconstitution inflammatory syndrome (IRIS) is playing a significant role.

Participants had been unwell for some time: a median 7 days. Published data on length of current illness in sepsis in both high and low-resource settings is lacking, but what data there are from elsewhere in sSA suggest that this is not unusual[9,25,129,296]. Barriers to accessing care were not addressed in this study and so the reasons for delaying hospital attendance (including the role of patient and healthcare factors) are not clear; 55% of participants had sought care for their current illness prior to presentation at the hospital, usually at the health centre. Optimum triage and other management of critically unwell patients at the health centre in a resource limited setting is not clear, and is likely to differ from hospital management. This could represent a fruitful area for future research.

The 28-day mortality of the cohort was 18% at 28 days, comparable to the pooled Sepsis-2 sepsis mortality from the systematic review and meta analysis presented in Chapter 1 (23% 95% CI [12-38%]) though considerably lower than the pooled Sepsis-2 severe sepsis mortality (49% 95% CI [39-58]). This is perhaps surprising as the inclusion criteria of this study include organ dysfunction criteria that are more similar to Sepsis-2 severe sepsis definitions than the Sepsis-2 sepsis definitions that are based on the systemic inflammatory response syndrome (SIRS), and would perhaps be expected to result in a higher mortality. In particular, the previous Malawian sepsis study (in the same hospital) from 2008/09 used a SIRS based definition of sepsis and found a mortality of 22%, with a severe sepsis mortality (defined using either 2 SIRS criteria and SBP < 90mmHg or any two of SBP < 90mmHg, capillary refill time > 2s, oxygen saturations < 90% or thrombocytopenia) of

50%[18]. There are several possible explanations for this. First, there is likely an effect of differing inclusion criteria: this study includes a respiratory rate criterion for recruitment, which has been shown elsewhere in sSA in large pooled datasets to be associated with mortality[297], a finding which was not replicated here. Second, sepsis mortality at our centre may be improving, either through improved management, or by population level changes resulting in improved health such as widespread ART coverage or reducing malaria incidence. Certainly, improving sepsis mortality in high-income settings is a trend that has been seen since the pivotal early goal directed therapy trial in 2001[5,85,122]. Third, participants in this study received reasonably intensive monitoring over the first 6 hours of their hospital attendance, which may have contributed to improved processes of care and hence improved outcomes. There is no way to address these hypotheses with the available data.

Participants continued to die post 28-days, to the end of the study period; this was most apparent in HIV-infected participants in whom there was a near-doubling of mortality from 19% at 28 days to 36% at 180 days. To my knowledge, this is the first data on post-30 day outcomes in sepsis in sSA, and demonstrates that longer term mortality following an admission with sepsis is a significant problem. The causes of late (post 28-day) death are unknown from this study. Given the advanced HIV of many of the participants, and the high prevalence of TB, opportunistic infection seems likely. Despite the suspicion of ART failure at baseline, switching to second line ART was unusual in participants who survived to discharge. This is perhaps not entirely surprising, as WHO and Malawian treatment guidelines mandate two elevated HIV viral load tests at least three months apart in the context of good adherence to diagnose ART failure and before switching to second line therapy, which may not have occurred during the six-month study period. Details of any HIV viral load testing performed as part of standard care (as well as any adherence counselling received) were also not captured in this study, so the proportion of true virologic ART failure are unknown. Nevertheless it is tempting to speculate that rapid adherence interventions or ART switching could improve post discharge outcomes in sepsis in Malawi. This could be an area of future research (Chapter 9), and is of increasing relevance as more people are exposed to first-line ART thanks to the success of global ART roll out.

In contrast to significant medium- to long-term mortality, health-related quality of life (HRQoL, as measured by EQ-5D-3L) seems to return to baseline by 12 weeks following sepsis admission, in contrast to high income settings where longer term morbidity is significant[89]. This may represent different patient populations with differing levels of physiologic reserve and capacity for recovery from critical illness, as the patient population in this study is significantly younger than a high-income setting sepsis population. It may also reflect the lack of resources available in LMIC: patients who would survive, but with disability, in a high-resource setting may die in a low-resource one. Nevertheless, the rapid return to a comparable HRQoL to healthy community controls following sepsis admission could make

improvement of sepsis outcome a cost-effective condition. Once again, to my knowledge, this HRQoL is the first available such data from sepsis in sSA, and can inform health economics analyses in sepsis here: the EQ-5D-3L utility scores can be used to calculate DALYs (disability-adjusted life years) in such an analysis.

### 3.5.2 Aetiology: TB dominates as a cause of sepsis

The aetiology of sepsis in this cohort is dominated by TB, with 34% of participants having at least one positive diagnostic test for TB. The majority of these were positive for urinary LAM. The prevalence of culture-confirmed MTB BSI was lower than expected; in previous studies of Sepsis-2 defined severe sepsis in Uganda[17,27] and Zambia[9,25] it was 28-40% in HIV-infected participants. In Malawi in the pre-ART era[99] the prevalence of MTB BSI in febrile adults at QECH was 14% (21/173), and 9% (9/104) in 2011 in the same centre in HIV-infected adults admitted with fever and chronic cough[106]. The 6% (8/138) I find here in HIV-infected participants therefore seems low. This could be due to technical (e.g. bottle under or over filling with blood) or laboratory factors, though the latter seems unlikely as the testing was carried out at the same laboratory and with the same SOP as the 2011 study by Feasey et al[106]. This could also be a true finding: given the association of MTB BSI with mortality the lower than expected mortality of this cohort could go hand-in-hand with a lower than expected MTB BSI prevalence, for example, or the high ART prevalence (despite the suspicion of common ART failure) could have an effect on the prevalence of MTB BSI.

Other identified causes of sepsis are as might be expected. *Salmonella* Typhi was the commonest blood stream infection isolate, reflecting the ongoing Typhoid epidemic in Blantyre which began in 2011[298], and seemed to be associated with HIV-uninfected participants, as has been previously described[299].

### 3.5.3 Determinants of mortality

Several expected markers of disease severity were associated with 28-day mortality, as have previously been described[7,18,27]. Unexpectedly, tachypnoea was associated with survival. The reasons for this are not clear. It could be that participants better able to mount an inflammatory response (and hence an increased respiratory rate) are more likely to survive, or that the effect is driven by low respiratory rates in the very unwell. It could also be explained by bias: elevated respiratory rate was an inclusion criterion for the study, as were shock, reduced conscious level and hypoxia. If unmeasured factors associated with mortality were also associated with these other inclusion criteria then this could cause an apparent mortality benefit to tachypnoea: a collider bias, in the language of causal inference.

A diagnosis of malaria was strongly associated with survival to 28 days. In Malawi - a malaria endemic country - most adults will have some level of immunity[300], which may explain this finding, but it is possible that the rapid diagnosis and treatment facilitated by the availability of malaria rapid diagnostic tests also contributes. Meningitis - all cryptococcal - strongly predicted death by 28 days, reflecting the well-described high mortality of this condition[301].

There was no signal of an association between time to antibacterials or volume of IV fluid administered and survival to 28 days. In high-resource settings, rapid administration of antimicrobials has been shown to be associated with improved survival in sepsis[6,7,126], and all sepsis guidelines stress the importance of rapid administration of antimicrobials[302]. This is based purely on observational evidence and no RCT has ever been (or will be, given the ethical issues) carried out; these studies are all open to confounding and require adjustment for disease severity. In this study, no significant effect of time-to-antibacterials was seen, though it is important not to interpret this lack of detected effect as lack of effect. The largest study to address this question, in a high income setting (New York, USA) found an adjusted odds ratio of 1.04 (95% CI 1.02-1.05) for death per hour delay of antibiotics, and included 49,331 participants[7], so lack of demonstration of effect here could be due to underpowering. I also have not, in this chapter, attempted to provide an estimate of the effect of time-to-antibacterials on mortality that is adjusted for confounding. Confounding could easily mask any apparent effect (i.e. sick patients given antibacterials quicker). I present an attempt to address this using modelling in the next chapter.

However it is important to be cognisant of the differences of this cohort to sepsis cohorts in high-income settings: presentation here was subacute, participants were often unwell for many days, and pathogens such as *S. Typhi* and TB would perhaps not be expected to cause such fulminant illness as the Gram-negative pathogens that often cause sepsis in high-income settings. In this context, rapidity of antimicrobial administration may not be such a critical determinant of survival. Further, most participants simply did not have sepsis caused by a pathogen that would be responsive to broad-spectrum antibacterials.

There was no detected benefit or harm associated with volume of intravenous fluid administered. How to safely administer intravenous fluid in sSA is unclear after RCTs in children[10] and adults[9] have shown harm to be associated with aggressive fluid resuscitation. Participants in this study received a comparable volume of fluid to the usual care arm of the Zambian RCT in adults[9], given that the trial was recruiting participants with shock: median 2.0L (IQR 1.0-2.5L) by 6 hours versus 1.5L (1.0 - 2.0L) in this study. The intervention arm of the RCT received 3.5L (2.7-4.0L). It may be that participants in this study did not receive enough fluid to be harmful; that there was insufficient variation in fluid exposure to detect an effect on mortality, that the current study is underpowered (particularly with the lower than expected mortality); that only a subset of the participants

in this study (i.e. those with shock) would benefit from fluid; or that the Zambian study population differs in some way and so response to fluid was different.

Perhaps the most striking finding is the strong association of receipt of TB therapy with survival. Care must be taken in interpreting this as cause and effect, because of the risks of confounding, and especially considering the limitations of the modelling undertaken in this chapter (see limitations). A protective effect of TB therapy in sepsis is plausible, however, from prior studies: autopsy studies show that TB is under diagnosed in HIV-infected patients who die in hospital[110]. The STAMP trial[109] found a mortality benefit in some a priori subgroups of a strategy of screen-and-treat with urinary LAM for all HIV-infected inpatients, suggesting a significant burden of undiagnosed TB, and prior sepsis cohorts in sSA have found TB as a common cause of sepsis. A retrospective study of 149 HIV infected adults with sepsis in Uganda[303], 55 of whom received anti-TB therapy, found an association between receipt of TB therapy and survival in Sepsis-2 severe sepsis (hazard ratio 0.32 95% CI 0.13-0.80 from Cox proportional hazard model) but not Sepsis-2 sepsis (hazard ratio 1.24 95% CI 0.53-2.90), but is hampered by its retrospective design.

What, then, is the role of TB therapy in sepsis in sSA? RCTs of empirical TB treatment have not previously been successful. The REMEMBER trial recruited outpatients with CD4 cell count below 50 cells  $\mu L^{-1}$  and randomised them to isoniazid preventative therapy or full TB therapy, and found no mortality benefit[304]. STASIS found no difference in mortality between a strategy of Xpert and urine LAM screening versus empiric TB therapy in outpatients with CD4 count below 100 cells  $\mu L^{-1}$ [305] and TB Fast Track found no mortality benefit in empiric therapy for outpatients with CD4 cell count below 150  $\mu L^{-1}$  if they were randomised to an algorithm that started TB therapy if they were assessed as high risk for TB using a combination of diagnostic tests (including urinary LAM) and clinical features (including BMI and haemoglobin)[306]. However all of these studies recruited ambulatory outpatients; it may be that inpatients have more disseminated TB, or a higher baseline risk of mortality. Empiric TB therapy for sepsis in a high-HIV/TB burden setting is a strategy that has never been assessed in an RCT.

The WHO provides guidance on empiric TB therapy in inpatients, however[111]: *"Hospitalised HIV-infected patients in high TB burden settings with cough and so-called 'danger signs' (fever  $> 39^{\circ}C$ , inability to stand, respiratory rate above  $30 \text{ min}^{-1}$ , heart rate above  $120 \text{ min}^{-1}$ ) should receive broad spectrum antimicrobials for 3-5 days, and, if there is no improvement, consider empiric TB therapy"*. This strategy was largely based on expert opinion, but has been shown to improve survival compared to usual care in a before-after study in South Africa[112]. Whether a 3-5 day delay will worsen outcomes in critically unwell patients with TB is unknown. There was no apparent relationship seen in this study between time to antitubercular therapy and death, but numbers were small ( $n=53$ ), and TB therapy administration was reasonably rapid, with a median of 120.6 hours

from admission to administration of TB therapy; 56% (35/53) of participants received TB therapy in less than 5 days.

In this context, the finding of a putative survival benefit for TB treatment in participants with sepsis is worth exploring further. Is it possible to produce unbiased estimates of the association of TB therapy with mortality, given the problems with the modelling approach presented in this chapter? Is it possible to move beyond association and assess the causal effect of TB therapy on mortality? Is any benefit confined to particular subgroups, especially groups that can be easily identified in low resource settings, to guide future sepsis treatment protocols? I take up these questions in the following chapter.

#### 3.5.4 Limitations

There are limitations to this study. There is no community control group, so it is not possible to say whether the positive malaria rapid tests in this study represent incidental parasitaemia or disease. Malaria films could perhaps inform this question by quantifying parasitaemia, but were not done. Only one aerobic blood culture and mycobacterial blood culture were done, and both tests are known to have suboptimal sensitivity with only a single culture[307,308]. No anaerobic culture was possible. HIV viral load testing was not done due to resource limitations.

There are several flaws with the logistic regression models used in this chapter that mean the parameter estimates from them may be biased. Separation was a significant challenge. This is a phenomenon where some predictor variable levels perfectly predict outcome, meaning that parameter value estimates become unstable. In this case, this occurred with malaria and meningitis, so these two predictor variables - both of which were very strongly associated with the outcome variable - were excluded from the model. If other variables were associated with these excluded variables then this could give biased estimates thanks to confounding. In addition, the model is very likely over-parametrised - around ten outcomes for each predictor variable are needed to avoid this[309] meaning that the out-of-sample predictive ability of this model would likely be poor, and would restrict the generalisability of findings. There is a risk that the model is fitting to noise rather than identifying the true data-generating process. Choosing which variables to include in the model, however, is not an easy task, and strategies such as stepwise variable inclusion have been shown to produce biased parameter estimates[309]. It may also be that some of the predictor variables are collinear - tachycardia and increased respiratory rate are associated with shock, for example, which can inflate the apparent standard errors of parameter estimates. Finally, it is likely that the mortality hazard of some variables included in the model is mediated by other variables (e.g. HIV mortality hazard may be disease mortality), and so interpretation of the parameters is difficult without an explicit causal framework[310].

### 3.6 Conclusions and further work

In conclusion, this chapter presents an in-depth clinical and microbiologic assessment of sepsis in Blantyre, Malawi, and finds that sepsis here is in some ways a subacute illness, with the dominant cause being tuberculosis. Nevertheless, long-term mortality is significant, and empiric TB therapy seemingly has a strong protective effect. Given the likelihood of confounding, arising partially from difficulties in the logistic regression modelling strategy I have used, this latter conclusion should remain speculative. In the next chapter, I extend the modelling presented here to attempt to address the problems I have identified.

Notwithstanding the next chapter of this thesis, some further work is planned.

- 49% of the cohort still have no diagnosis, and further testing for leptospirosis, rickettsioses and arboviruses in the first instance, using acute and convalescent sera is underway.
- The reason for the low prevalence of TB BSI in combination with a high prevalence of urine LAM positivity is unknown, and running Xpert ultra on stored blood samples may help to understand if there were technical problems with the mycobacterial blood cultures.
- Fluid administration in sepsis in sSA is clearly complex, and longitudinal modelling of response to fluid over the first six hours of hospital admission in this cohort is planned.



## Chapter 4

# Modelling to identify determinants of sepsis mortality

### 4.1 Chapter overview

In this chapter I develop the mortality models from Chapter 3 to address difficulties arising from the problems of separation, overparametrisation, collinearity and missing data. I use Bayesian logistic regression following multiple imputation of missing data and dimensionality reduction using factor analysis of mixed data (FAMD), to show that, broadly, inferences from the original models are sound: well recognised markers of severity were associated with death. In terms of treatment determinants of outcome, the association of receipt of TB therapy with survival persisted across all models. In contrast to high-income settings, there was no clear association between more rapid administration of antibacterials or volume of intravenous fluid administered and survival, though in both cases the 95% credible intervals of effect size incorporated a clinically significant effect. A subgroup analysis using a propensity-score approach suggested that the association of TB therapy with survival was strongest in immunosuppressed and/or anaemic participants, though numbers were small and confidence intervals wide. The role of early administration of TB therapy in septic adults in Malawi is unknown, though the analyses here suggest a potentially significant impact, and, I suggest, contribute to the equipoise necessary for clinical trials.

### 4.2 Introduction and chapter aims

In Chapter 3, I used classical epidemiological approaches to identify associations with mortality in sepsis in Malawi, however there were some problems with the approaches used, both technical and conceptual. The technical issues are: that the model may include too

many parameters and be overfit; that a complete-case analysis was undertaken which can introduce bias; that included variables may be collinear which can contribute to increased variance in parameter estimates; and, perhaps most importantly, some parameters perfectly predicted outcome and so were excluded from the model. Conceptually, I argue that it is difficult to interpret parameters without first making explicit the causal model for data generation. I will cover each of these points below, before describing the techniques I used to address them.

First, the model may be overfit. Adding covariates to a model may reduce bias in parameter estimates but will increase variance (the so called bias-variance trade off[311]) which may make interpretation difficult. More parameters in a model results in a better fit to the data, but an overfit model may end up fitting to noise, rather than identifying the true data-generating process, and so would have biased inferences and poor out of sample prediction. Rules of thumb based on  $\sim 10$  outcomes per included predictor have been suggested[309], but the process of selecting variables to include is difficult with no consensus as to how it should be achieved. Common approaches include stepwise inclusion strategies, where variables are sequentially added or removed based on some criteria of model fit or statistical significance, but these can introduce significant bias. This is because the statistics used to test the parameters (and generate confidence intervals around effect sizes) are based on an assumption that a single hypothesis is being tested, an assumption which is violated by the stepwise model building process. It can be shown that standard errors are too small, that p-values are biased towards zero and parameter estimates biased away from zero[309].

How, then, to select variables to include in the models presented here? *A priori* selection of variables for theoretical reasons is likely ideal, but this becomes difficult when there are a large number of potentially important predictors, as variance in parameter estimates may be increased. Dimensionality reduction techniques (such as principal components analysis) or shrinkage methods (lasso or ridge regression) have been suggested as alternative predictor variable selection techniques[309,311,312]. I use a dimensionality reduction technique called factor analysis of mixed data[313] (FAMD) and compare out-of sample prediction (using cross validation) of models built using this technique to the original model - see methods, below.

A further problem in modelling mortality in studies of sick inpatients is collinearity, where some predictor variables can be predicted with high accuracy by other predictor variables. For example, shocked patients are likely to have elevated lactate, low blood pressure, low bicarbonate, and high heart rate and so uncertainty in parameter estimates becomes very large (or the parameters are unidentifiable) when these are all entered a regression model together[314]. An advantage of principal-components type dimensionality reduction (including FAMD) is that they can solve this problem by generation new coordinate systems that are constrained to be orthogonal. Missing data bias too can be significant; I address

this by using multiple imputation by chained equations where new values for each missing data point are generated by models based on all other data[315].

The final technical modelling challenge to overcome is the phenomenon of separation, where some covariates perfectly predict outcome[314]. In the maximum likelihood framework (the method usually used to fit logistic regression models) the parameter estimates for such covariates are non identifiable (essentially infinite), and hence the covariates are often excluded from the model, even though they are often very strong predictors of outcome. I use Bayesian logistic regression with weakly informative priors to overcome this problem. In the Bayesian modelling framework, probabilities encode our belief about parameter values, rather than encoding a long-term expectation of parameter estimates following multiple samples from a population, as in the frequentist framework. Probability distribution of parameters are generated from a combination of the data and a *prior* probability distribution, quantifying our belief about the parameter values prior to seeing the data. To overcome the problem of separation, Gelman et al [316] suggest using our belief that very large or very small parameter estimates (e.g. an odds ratio of 100) are unusual, which is encoded as a prior with a student's t distribution with 3 degrees of freedom, centre zero and scale 2.5. In the broadest sense, this is a distribution not dissimilar to a normal distribution, but with longer tails: we think that the parameters will be close to 0, but allow a chance they may be larger. In effect, this pulls the infinite parameter estimates from the data closer to 0.

Even if the parameters of the regression can be correctly specified, however, correct interpretation of predictor effects is often difficult or impossible without a clear hypothesised causal structure. For example, consider a hypothesised causal structure of death in sepsis in Figure 4.1, which I express as a directed acyclic graph (DAG); nodes represent collections of variables which theoretically specify host status (age, sex, immune status including HIV status and CD4 cell count), infection type (e.g. causative pathogen, site), disease severity (e.g physiological variables quantifying shock, hypoxia etc.), therapies administered, and outcome. Arrows (called edges in the DAG framework) show causality: host status influences infection (e.g. TB is more common in HIV) and severity (patients with advanced HIV may have more severe infection), for example, and therapies administered is likely to be influenced by disease severity (perhaps sicker patients receive antimicrobials more quickly), host status (clinicians are likely to administer different therapies to HIV-infected patients), and infection type. A standard analysis of sepsis would construct a predictive multivariable model for death by including factors which the analyst felt likely to be associated with mortality, which would usually include HIV status, CD4 cell count, physiologic variables (such as presence of shock) and infection variables (e.g. presence of bloodstream infection [BSI]). The effects of the predictor variables are often then interpreted as the independent effect of the included predictors, after controlling for all others; however, this may not be the case.

For example, severity is at least in part a mediator of the effect of HIV on outcome, so the interpretation of the coefficient of HIV in such a model is the residual effect of HIV once disease severity is accounted for. It is likely that there are direct effects of host and infection factors on outcome (dotted edges in Figure 4.1), not least because measured variables in a study are unlikely to wholly quantify disease severity, but if not then controlling for disease severity will completely remove the effect of HIV status on mortality, which may not be the analysts intention, or interpretation of parameters. This has been called the “Table 2 fallacy”[310]. It is important therefore to clearly define the effect that is being sought from an analysis (e.g. the effect of HIV status on mortality) and to ascertain which factors need to be controlled for based on this. It may be that a number of different models are necessary to estimate parameters of interest, if more than one parameter is of interest. The causal inference framework provides tools to do this using DAGs[317], and the *dagitty* package in R[318] automates this framework so, when provided with a DAG, it can output the variables that must be conditioned upon to estimate the causal effect of an exposure on an outcome. In this chapter, therefore, I am clear that the aim of the analysis is to provide an estimate of the effect of treatments administered on mortality; the class of antimicrobial administered (antibacterial, antifungal, antimycobacterial or antimalarial) as well as the time-to-antimicrobial for different classes, and the volumes of intravenous fluid administered. This will inform the overarching aim of the thesis - to develop novel antimicrobial strategies for sepsis in sSA to improve outcomes.

### 4.3 Methods

Assuming the causal model in Figure 4.1, an estimation of the effect of administered treatment will require correcting for (or conditioning on) host, infection, and severity variables (assuming a direct effect of infection and host on outcome, as seems likely) i.e. all the variables that were included in the logistic regression model in Chapter 3. To solve the problem of nonidentifiability of the models including malaria and meningitis status, I refit the models in a Bayesian framework with weakly informative priors. All continuous variables were scaled to have mean 0 and standard deviation 1, and a student’s t distribution centred on 0 with three degrees of freedom and a scale of 2.5 was used as the prior for all parameters, following Gelman et al[316]. The model was fit using the *brms* package in R[319], which acts as a front end to the Stan probabilistic programming language[320]. Four Markov-chain Monte-Carlo (MCMC) chains each with 1000 iterations and a burn-in of 500 iterations were used with default *brms* settings. Convergence was assessed using traceplots and assessing for autocorrelation using the Gelman-Rubin diagnostic ( $\hat{R}$ ) with a target of  $\hat{R} < 1.1$ . Parameter estimates were expressed as medians and 95% credible intervals.

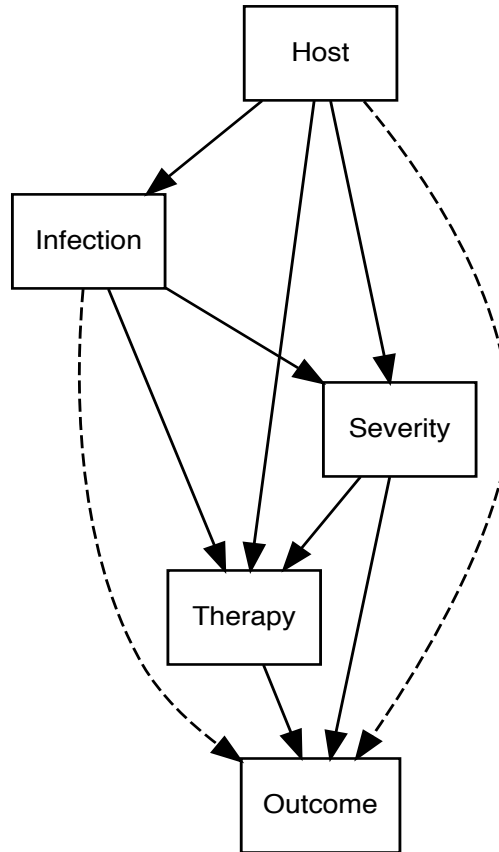


Figure 4.1: Hypothesised causal structure of death in sepsis. Host factors (e.g. age, sex, immune status) influence the type of infection; disseminated TB is more common in HIV, for example. Severity (variables quantifying e.g. shock or respiratory failure) is influenced by infection type and host factors. Therapy encodes which antimicrobials were administered and rapidity of administration of antimicrobials, and is influenced by disease severity (sicker patients may be given different therapies), host factors (HIV status may influence treatment) and the infection type (for example, malaria rapid diagnostic tests influencing rapidity of malaria treatment). Dotted edges from host and infection to outcome are because it is not clear *a priori* whether the effect of infection and host factors are entirely mediated by disease severity: in fact, even if this were the case in a theoretical sense, the available severity variables are unlikely to completely account for the causative effect of infection type on mortality and so conditioning on all available severity variables is likely to leave some residual causative effect of infection type. See text for further discussion.

To correct for missing-data bias, missing data were imputed using multiple imputation of chained equations using default settings in the *mice* package in R[321], with each missing variable predicted by all other missing variables in the model, to produce 5 imputed datasets. Models were fit using *brms* and then pooled parameter values calculated by taking medians and 95% confidence intervals of pooled posterior parameter estimates from all imputed datasets. The parameter estimates from complete-case analysis and multiple imputation are both presented.

One of the concerns of this model is that it is overfit - that is, there are so many parameters that it will fit to noise in the data rather than to the true data-generating process. To assess whether this was the case, I performed dimensionality reduction to collapse the predictor variables into a smaller number of variables, refit models using these variables, and then compared all models predictive ability using leave-one out cross validation. The dimensionality reduction technique that I used was factor analysis of mixed data (FAMD) from the *FactoMineR* package in R[313]. This technique uses principal component analysis (PCA) for continuous variables and multiple correspondence analysis (MCA) to generate a new orthogonal coordinate system which maximises explained variance in each FAMD axis. FAMD axis one therefore explains the most variance in the dataset, followed by FAMD axes 2 and 3, and so on. As well as reducing the dimensionality of the dataset, this technique has the advantage of ensuring an orthogonal coordinate system to tackle the problem of collinearity. The raw covariate values were used to generate these new coordinate system.

Because the exposures of interest are the therapies administered to the participants in the study, treatment variables (receipt of antibacterials, antifungals, antimalarials, antimycobacterials, and IV fluid) were left untransformed. These variables and a number of transformed FAMD variables (ranging from 1 to 5) were used as predictors in new models to predict death by 28 days. The out of sample predictive ability of the models was assessed by performing leave-one-out cross validation using the *loo* package in R[322]. This estimates the out-of sample predictive ability of the model by estimating a quantity called the expected log predictive density (*ELPD*) essentially the logarithm of the likelihood for a new, unseen dataset conditional on the current data. This quantity is estimated using leave-one-out cross validation to produce an estimate of the *ELPD*, hereafter referred to as  $ELPD_{loo}$ . The standard error of  $ELPD_{loo}$  for a model is also calculated and so two models can be compared by comparing the  $ELPD_{loo}$  difference and standard error; if the difference is greater than twice the standard error (i.e. a 95% confidence interval, assuming normality) we can be confident that one model would be expected to have greater out-of-sample predictive ability than the other[322].

The relationship between time-to-antimicrobials and mortality was assessed, initially in bivariate associations using nonparametric locally estimated scatterplot smoothing (LOESS) regression which performs a rolling linear regression[323] and estimates the probability of

death by 28 days as a function of the predictor variables. Only for antibacterials were there sufficient data to construct regression models which used time to antibacterial therapy as a predictor for death by 28 days, alongside the other treatment variables and the first three FAMD dimensions. In view of possible nonlinear relationship between time to therapy and death apparent in the bivariate plots both linear and second-order polynomial models were fit. Coefficient estimates are presented, but because interpretation of polynomial coefficients is challenging, predicted probability plots with 95% credible intervals with the levels of the other covariates set to their mean values were plotted, using all the posterior draws to generate the median prediction and 95% credible intervals.

Finally, to attempt to correct for confounding using a different method, a propensity-score matching approach was used to produce an unbiased estimate of the effect of receipt of TB therapy on 28-day mortality. Variables that had been identified as being associated with mortality from the models described above, along with variables that were associated with receipt of TB therapy apparent on bivariable analysis were included in a (maximum-likelihood fit) logistic regression model to generate a propensity score. Because HIV-uninfected participants did not have a CD4 count measured, a new dichotomous variable was used which was coded as 1 for HIV-infected participants with a CD4 count below 100 cells  $\mu L^{-1}$  and 0 for everyone else. The variables used to generate the propensity score were therefore: haemoglobin, HIV status, CD4 count, respiratory rate, oxygen saturations, inability to stand, diagnosis of BSI, diagnosis of TB, diagnosis of meningitis, diagnosis of malaria, bicarbonate, and heart rate. Participants were then matched 1:1 on this propensity score without calliper restriction using the *MatchIt* package in R[324] and the distribution of covariates in this new cohort examined using kernel density plots. Effect of TB therapy on mortality was then expressed as risk ratios, and subgroup analysis carried out to explore whether there was any effect modification of the apparent effect of TB therapy in advanced immunosuppression (defined as CD4 cell count below 100  $\mu L^{-1}$ ), or anaemia (defined as haemoglobin below 8g  $dL^{-1}$ ) or confirmed TB.

## 4.4 Results

Bayesian logistic regression with weakly informative priors succeeding in fitting the models from Chapter 3, where a maximum likelihood approach had failed; the inferences - particularly concerning the apparent association between TB therapy and survival - were largely unchanged, including after multiple imputation of missing data (Table 4.1). Because of concerns about overfitting, dimensionality reduction with FAMD was carried out; the first four FAMD dimensions explained 42% of the variance in the dataset, a not inconsiderable amount for 21 predictor variables. The composition of FAMD dimensions one, two, three and four are shown in Figure 4.2. FAMD dimensions 1 (explaining 14% of the total variance

of the dataset) was composed primarily of variables that were associated with shock (urea, lactate, low MAP and bicarbonate, and creatinine) and HIV-associated immunosuppression (HIV status and haemoglobin); FAMD dimension 2 was similar but included organ dysfunction criteria that were orthogonal to FAMD 1: low oxygen saturation and platelet count, and male sex. FAMD 3 was composed of absence of malaria, low GCS, high white cell count and inability to stand; FAMD 4 with young age, fever, and tachycardia.

Table 4.1: Unadjusted and adjusted odds ratios of death by 28 days in sepsis from Bayesian logistic regression, for complete case analysis (CCA) and following multiple imputation of missing data.

Variable	aOR (95% CrI)	
	CCA	Imputed
<b>Host Variables</b>		
Age (per 5 years increase)	0.92 (0.66-1.25)	0.87 (0.65-1.14)
Male sex (vs female)	0.91 (0.22-3.51)	0.67 (0.18-2.36)
HIV Infected (vs uninfected)	0.21 (0.03-1.22)	0.32 (0.06-1.59)
Haemoglobin (per g dL <sup>-1</sup> )	<b>0.71 (0.54-0.91)</b>	<b>0.69 (0.52-0.90)</b>
<b>Severity Variables</b>		
Temperature ( per °C)	0.72 (0.30-1.73)	0.56 (0.26-1.18)
Heart rate (per 10 min <sup>-1</sup> )	1.20 (0.87-1.72)	1.14 (0.85-1.57)
Mean arterial BP (per 10 mmHg)	1.20 (0.73-1.93)	1.14 (0.70-1.80)
Respiratory rate (per 10 min <sup>-1</sup> )	<b>0.25 (0.08-0.66)</b>	<b>0.38 (0.16-0.88)</b>
Oxygen saturation (per 5%)	0.73 (0.48-1.11)	<b>0.67 (0.45-0.99)</b>
GCS (per 1 unit)	0.76 (0.50-1.12)	0.75 (0.51-1.10)
Unable to stand	<b>13.79 (2.88-74.50)</b>	<b>13.64 (3.35-64.82)</b>
Lactate (per 1 mmol L <sup>-1</sup> )	1.12 (0.91-1.39)	1.13 (0.92-1.41)
White cell count (per 1x10 <sup>9</sup> L <sup>-1</sup> )	0.96 (0.84-1.07)	0.94 (0.83-1.05)
Platelet count (per 100x10 <sup>9</sup> L <sup>-1</sup> )	1.13 (0.70-1.78)	0.94 (0.60-1.43)
Bicarbonate (per 1 mmol L <sup>-1</sup> )	0.97 (0.81-1.17)	0.95 (0.81-1.11)
Urea (per 1 mmol L <sup>-1</sup> )	1.20 (1.00-1.45)	1.17 (1.00-1.37)
Creatinine (per 10 mmol L <sup>-1</sup> )	0.99 (0.92-1.08)	0.99 (0.93-1.08)
<b>Diagnosis</b>		
BSI (vs no BSI)	<b>0.04 (0.00-0.48)</b>	<b>0.04 (0.00-0.40)</b>
TB (vs no TB)	1.12 (0.25-5.00)	0.72 (0.18-2.69)
Malaria (vs no malaria)	0.01 (0.00-2.27)	<b>0.00 (0.00-0.41)</b>
Meningitis (vs no meningitis)	<b>68.53 (1.29-27384.82)</b>	<b>37.00 (1.03-6237.92)</b>
<b>Treatment Received</b>		
Received antibacterial (vs none)	8.38 (0.20-6631.38)	1.46 (0.10-30.00)
Received antifungal (vs none)	1.39 (0.23-8.60)	1.19 (0.24-5.65)
Received antimalarial (vs none)	0.03 (0.00-8.68)	0.08 (0.00-13.22)
Received antimycobacterial (vs none)	<b>0.11 (0.02-0.58)</b>	<b>0.12 (0.02-0.56)</b>
IV fluid (per L)	0.82 (0.29-2.21)	0.79 (0.31-1.98)

*Note:*

CCA = Complete case analysis, BP = Blood pressure, GCS = Glasgow coma scale, BSI = Bloodstream infection, TB = tuberculosis. All odds ratios are for as increase in the variables shown.

Graphically, FAMD dimensions one (and perhaps two) appeared to show an association with mortality (Figure 4.3), which was confirmed on logistic regression modelling. The first 5 FAMD dimensions were used to fit models predictive of death by 28 days, along with untransformed treatment variables; the primary interest here was to see if the apparent effect of treatment administered would change under these models, but a secondary effect



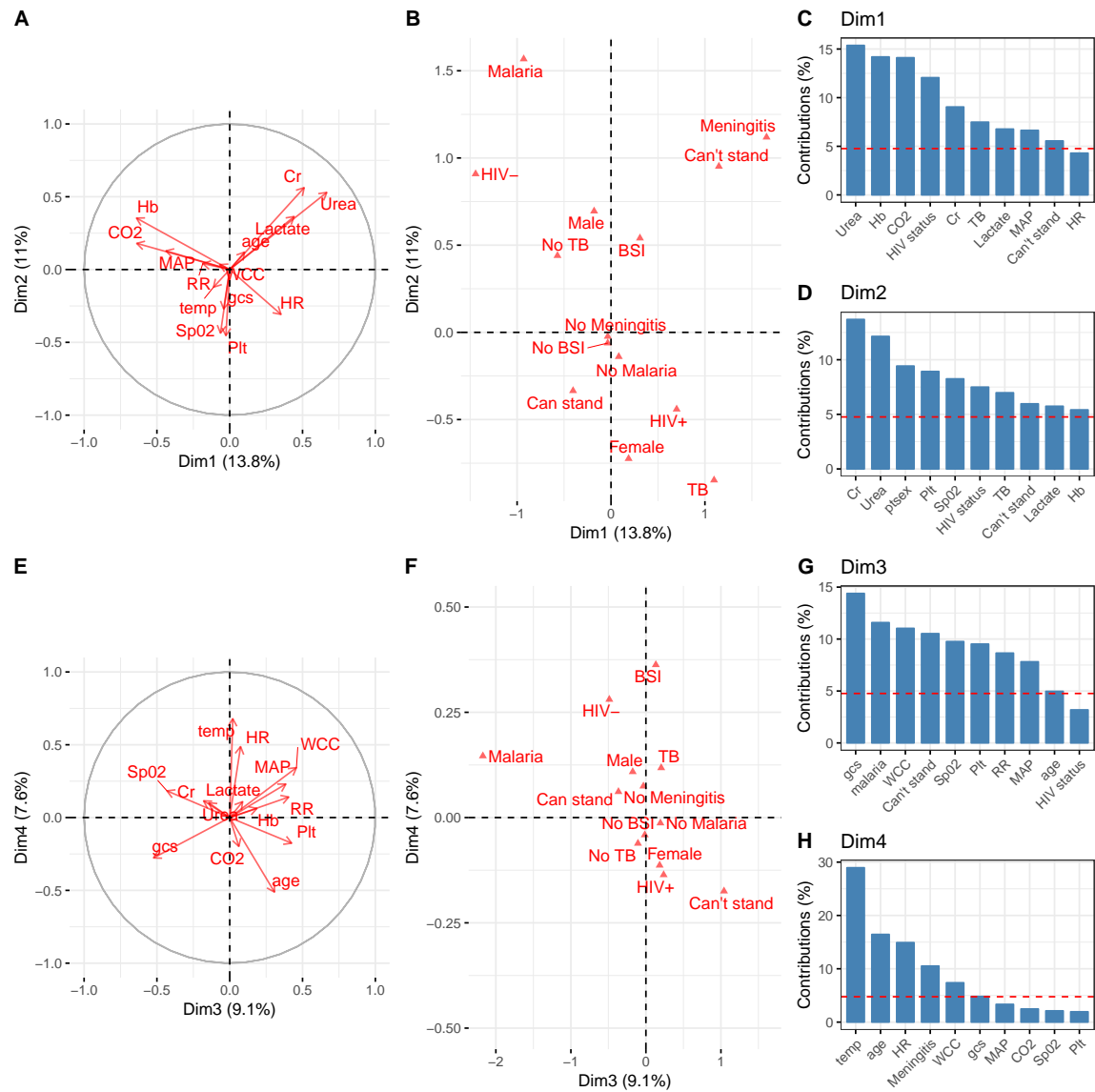


Figure 4.2: Composition of FAMD variables 1-4. FAMD uses both principal coordinate analysis (PCA) and multiple correspondence analysis (MCA) to generate new orthogonal coordinates. A and E show projections of the original variables into the new PCA space to show, for example, that a higher haemoglobin is associated with smaller (more negative) values of FAMD dimension 1. B and F show the positions of the categorical variable values in MCA coordinate space. C-D and G-H show the composition of the FAMD dimensions 1-4 in terms of the untransformed variables. Cr = serum creatinine, CO2 = serum bicarbonate, gcs = Glasgow coma scale, Hb = haemoglobin, HR = heart rate, Malaria = diagnosis of malaria, MAP = mean arterial blood pressure, Meningitis = diagnosis of meningitis, Plt = platelet count, ptsex = sex, RR = respiratory rate, SpO2 = capillary oxygen saturation, temp = temperature, TB = diagnosis of TB, WCC = white cell count.

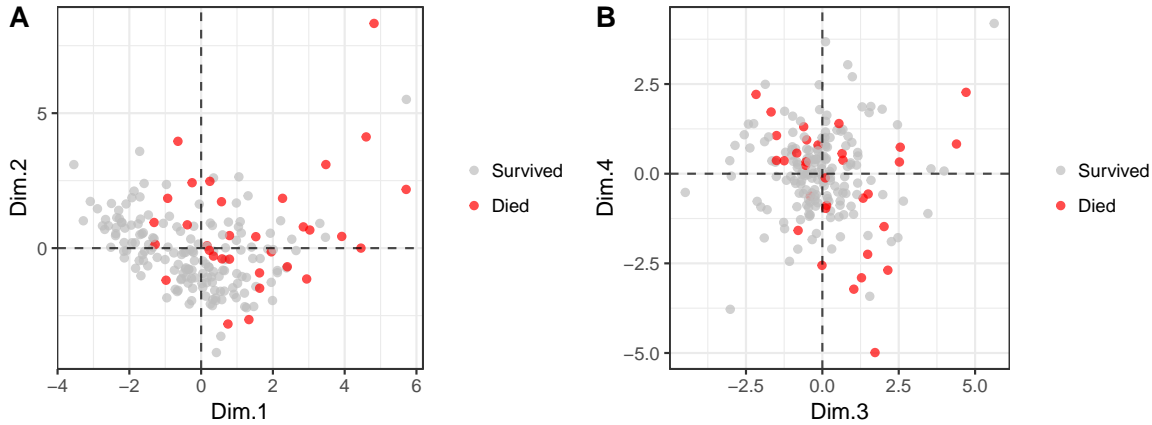


Figure 4.3: Individuals plotted in FAMD space projected onto FAMD dimensions 1 and 2 (A) or 3 and 4 (B) with patients who died by 28 days coloured red. There is an apparent correlation between higher values of FAMD dimensions 1 and perhaps 2 with death (more red points in the upper right quadrant)

was to provide estimates of association of FAMD dimensions with death. Five models were fit using one, two, three, four or five FAMD dimensions; parameter estimates from these models are shown in Figure 4.4A, along with the parameter estimates from the original model using all, untransformed, parameters. Parameter estimates from treatment variables were largely unchanged across all models, though uncertainty was markedly increased in the original models. Nevertheless, inferences were largely unchanged: we can be confident only that the odds ratio of the effect of TB treatment is different to zero. FAMD dimensions one was strongly associated with mortality; FAMD dimensions two and three less so. FAMD dimensions 4 and 5 showed no convincing association; in fact, FAMD dimension 4 seemed, if anything, protective, though 95% credible intervals of odds ratios crossed 1.

The out-of-sample predictive ability of the models was assessed using the expected log predictive density (ELPD) estimate from leave one out cross validation. In absolute terms, all FAMD models greater ELPD than the original model but any differences were small compared to the standard error of the ELPD estimate. We can not be confident that any model has different out of sample predictive accuracy and therefore can be as confident in the parameter estimates from the original (untransformed) model as any other.

#### 4.4.1 Exploring time-to antibacterials and IV fluid as determinants of mortality

Exploration of bivariate associations of mortality with time to antimicrobials and volume of intravenous fluid received are shown in Figure 4.5, where LOESS moving linear regression provides a nonparametric estimate of probability of death by 28 days as a function of treatment variables. Time to antimalarial therapy is not shown in this plot as no patient who received antimalarial therapy died. Volume of intravenous fluid administered has no

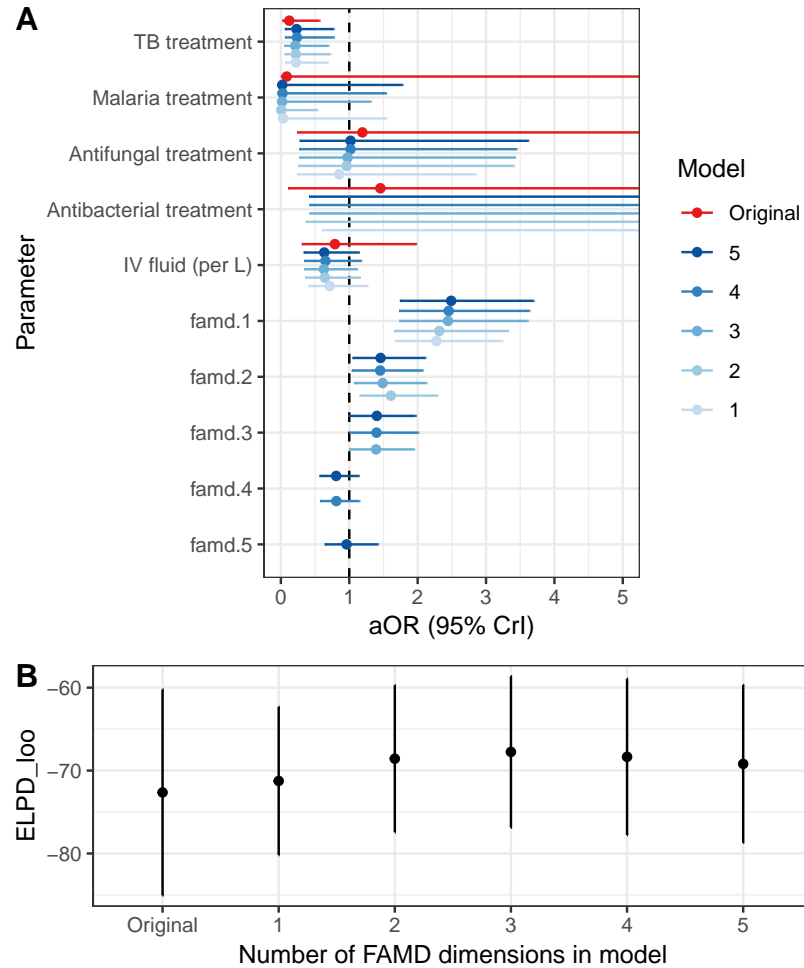


Figure 4.4: Modelling the effect of receipt of different treatments following dimensionality reduction with factor analysis of mixed data (FAMD). A (Top) shows parameter estimates for treatment variables only from the original imputed model using all raw covariate values. Models 1-5 use the first 1,2,3,4 or 5 transformed dimensions from FAMD. Parameter estimates and inferences are essentially unchanged, though there is less uncertainty generally in the estimates from the FAMD models. This would be expected as fewer parameters with less collinearity, are used. B (bottom) shows the estimated ELPD (expected log predictive density) from leave-one out cross validation from all the models, along with the standard error of the estimate. This is a measure of out of sample predictive accuracy: bigger (less negative) is better. One of the concerns of the original model is that it is overfit and so would have poor ELPD. In absolute terms this is true but the magnitude of the difference is much less than the standard error, meaning that out of sample prediction for all the models is broadly similar, giving confidence in the original model inferences.

apparent effect on 28 day mortality (Figure 4.5A). It might be expected that any effect would be most apparent in participants with shock: stratifying the analysis by shock (defined as mean arterial blood pressure below 75mmHg, Figure 4.5B) once again showed no apparent relationship. Neither time to antimycobacterial or antifungal therapy showed any apparent association though confidence intervals are wide (Figures 4.5C and D).

There was no apparent relationship between time to antibacterials and 28 day mortality up to around 40 hours, when there was a suggestion of an increased probability of death (Figure 4.5E). To explore this further, I used a logistic regression analysis, including only patients who received antibacterials ( $n = 207$ ) using both linear models, fitted in a Bayesian framework as before (and following imputation of missing data), and, in view of a possible nonlinear effect, second order polynomial models. Model three from the analysis above was used (incorporating the first three FAMD dimensions), as the best fitting model. The estimates of the coefficients of the linear model is shown in Table 4.2 and the predicted probability of death by 28 days shown in Figure 4.5. In both cases it is not possible to fully rule in or out an effect of antibacterial delay. The 95% credible interval of the adjusted odds ratio for death per hour of antibacterial delay from the linear model crossed one (aOR 1.01 95% [CrI 0.98-1.04]) though incorporated a clinically relevant effect size, and the uncertainty in predictions from the polynomial of a late nonlinear effect of antibacterial delay are so wide that it is not possible to draw any conclusions.

Table 4.2: Adjusted odds ratio of death by 28 days per hour delay in antibacterials

Variable	aOR (95% CrI)
Time to antibacterials (per hour)	1.01 (0.98-1.04)
IV fluid (per L)	0.65 (0.36-1.16)
Received antimalarial (vs none)	0.02 (0.00-1.37)
Received antifungal (vs none)	1.03 (0.28-3.49)
Received antimycobacterial (vs none)	0.18 (0.05-0.61)
famd.1	2.56 (1.81-3.73)
famd.2	1.43 (1.04-2.01)
famd.3	1.38 (1.00-1.90)

*Note:*

The variables famd1,2 and 3 are the three transformed dimensions following dimensionality reduction using factor analysis of mixed data that account for the most variability in the dataset.

#### 4.4.2 Propensity score matching and subgroup analysis

Finally, I used propensity score matching, a different method to attempt to generate unbiased estimates of the effect of receipt of TB therapy on mortality. First I examined bivariate associations of receipt of TB therapy (Table 4.3 in the chapter appendix). Patients who received TB therapy were almost all HIV-infected (88% [46/52] vs 60% [95/161], between group difference 28% [95% CI 15-41%]) with lower CD4 count (median 60 vs 123

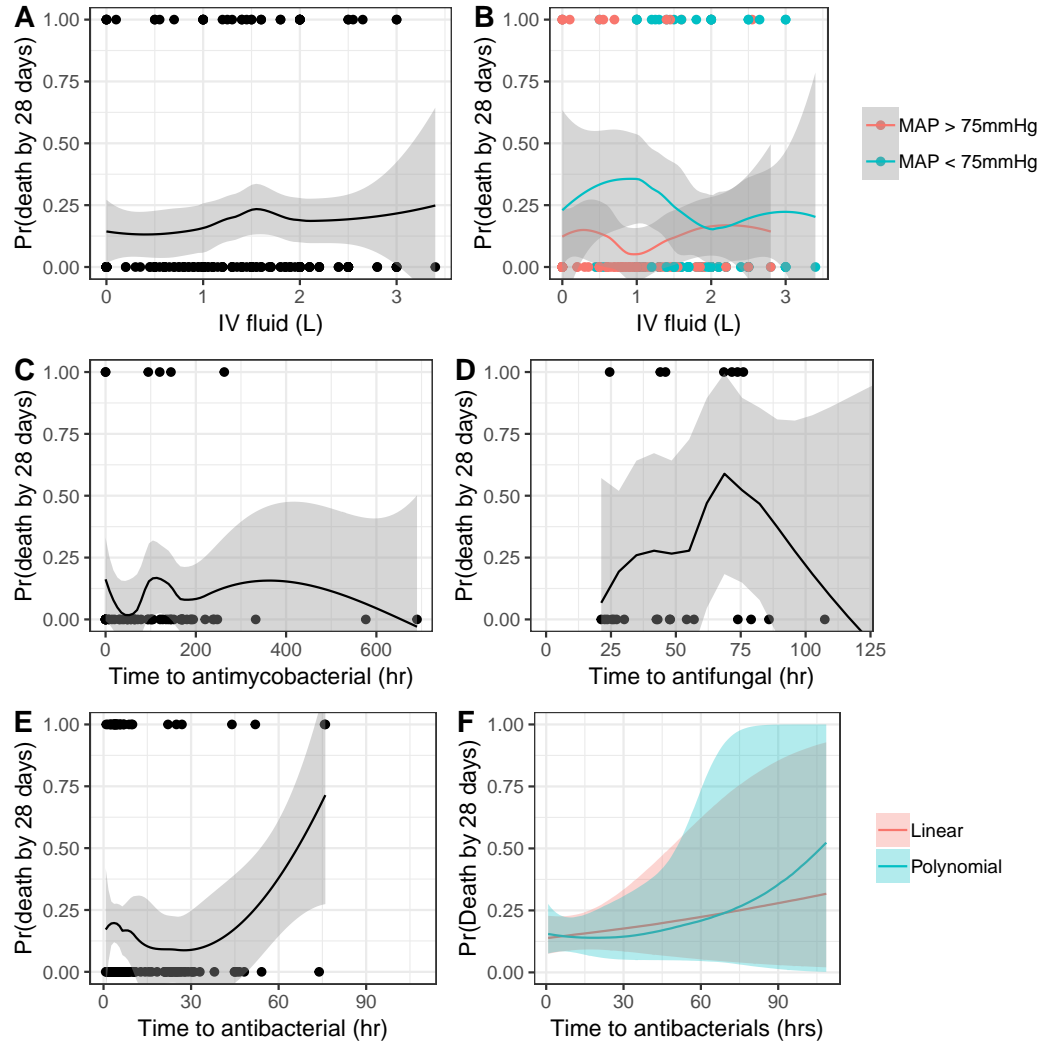


Figure 4.5: Associations of IV fluid volume and time-to-antimicrobials with death by 28 days. A-E show nonparametric regression (LOESS) of outcome (with death coded as 1 for died and 0 for survived) against various covariates; the regression line can be interpreted as the probability of death by 28 days and can be used to assess for a bivariate relationship and also the nature of any relationship (i.e. linear versus nonlinear). A: IV fluid (L), B: IV fluid stratified by presence or absence of shock (defined as MAP < 75mmHg), C: Time to antimycobacterials, D: Time to antifungals E: Time to antibacterials, with a possible late, nonlinear relationship. F: Models of time-to-antimicrobials as a predictor of mortality considering time-to-antibacterials to have a linear or second order polynomial effect. In both cases the uncertainty in the effect is such that there is no convincing relationship. Overall, there is no convincing relationship between any of these variables and death by 28 days.

cells  $\mu\text{L}^{-1}$ , between group difference 63 [15-142] cells  $\mu\text{L}^{-1}$ ) and haemoglobin (median 9.7 vs 11.1 g  $\text{dL}^{-1}$ , between group difference 1.4 [95% CI 0.5-2.9] g  $\text{dL}^{-1}$ ), with higher heart rate (median 125 vs 120  $\text{min}^{-1}$  difference 5 (1-13)  $\text{min}^{-1}$ ) and lower bicarbonate (18 vs 20  $\text{mmol L}^{-1}$  difference 2 (0-3)  $\text{mmol}^{-1}$ ) though most of these associations would be expected to pull an estimate of the mortality effect of TB therapy towards the null, rather than inflate an effect size. More patients with a positive diagnostic test for TB received TB therapy, as might be expected (53% [28/53] of those receiving TB therapy had a positive diagnostic test for TB, versus 28% [48/172] not receiving therapy, difference 25% (95% CI 9-41%)), though almost all the TB treatment was empiric, as the treating clinicians did not have access to urinary LAM results (which were batch processed on frozen urines) or mycobacterial blood culture results (which take up to 6 weeks to become positive).

Factors associated with receipt of TB therapy on univariate analysis (HIV status, CD4 count, haemoglobin, diagnosis of TB, heart rate, and bicarbonate) and factors associated with mortality from the models presented above (haemoglobin, respiratory rate, oxygen saturation, inability to stand, bloodstream infection and diagnosis of malaria) were used as predictors in a logistic regression to predict receipt of TB therapy. Predictions from this model were used to generate a propensity score for each participant, and then each participant who received TB therapy was matched with one participant who did not to generate a new cohort, with better matching of covariates (Figure 4.7). The propensity-score adjusted risk ratio of survival to 28 days in this cohort upon receipt of TB therapy was 1.22 (95% CI 1.03-1.46), similar to the unadjusted estimate (Figure 4.6). Mortality benefit seemed higher in the immunosuppressed and anaemic in absolute terms, though with significant uncertainty in the estimates (Figure 4.6): RR 1.59 (95% CI 1.02-2.50) in those with haemoglobin below 8g  $\text{dL}^{-1}$  compared to 1.09 (95% CI 0.92-1.30) above 8g  $\text{dL}^{-1}$ ; 1.42 (95% CI 1.04-1.93) in those with a CD4 cell count below 100 cells  $\mu\text{L}^{-1}$  compared to 1.04 (95% CI 0.89-1.22) above. Mortality benefit was also higher in those with a confirmed diagnosis of TB: RR 1.30 (95% CI 1.00-1.70) in those with confirmed TB compared to 1.13 (95% CI 0.90-1.43) without.

## 4.5 Discussion

I have presented, in this chapter, an analysis using a number of statistical techniques to further explore the associations of mortality in this cohort, and to address the difficulties encountered with the modelling strategies presented in Chapter 3. In addition I have presented an explicit hypothesised causal structure of death to facilitate a causal interpretation of the model parameters. In Chapter 3, maximum likelihood fitting of logistic regression models was potentially biased because some predictors were perfectly associated with outcome, and were excluded - the phenomenon of separation. This was overcome

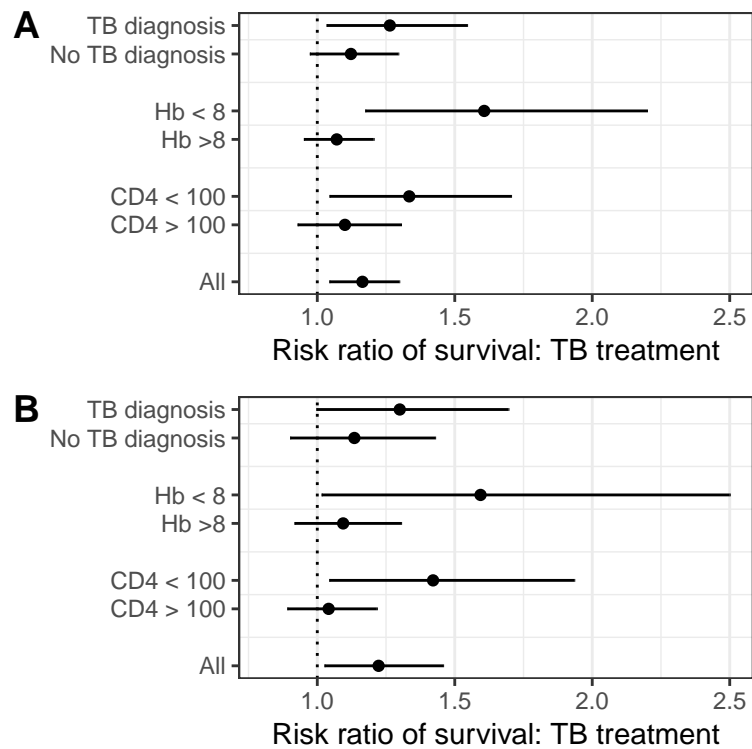


Figure 4.6: Subgroup analysis of effect of TB therapy on mortality. A (Top) shows crude (unadjusted) risk ratio for survival to 28 days;  $RR > 1$  favours TB therapy,  $RR < 1$  favours no TB therapy. A significant effect is seen in the immunosuppressed, anaemic, and to a lesser extent, those with a confirmed diagnosis of TB. B (Bottom) shows the same analysis for the propensity-score matched cohort, showing that the overall and subgroup effects are essentially unchanged, though uncertainty in parameter values is greater.

using Bayesian logistic regression with weakly informative priors. As expected, malaria and meningitis were strongly associated with survival and death, respectively, but more importantly inferences from the model were largely unchanged once these strong predictors were included in the model. The strong association of TB therapy with survival persisted. Unexpectedly, bloodstream infection showed a protective effect; given *S. Typhi* was the commonest bloodstream infection isolate, which is associated with a low mortality in the Blantyre context (2.5% since 2011[325]), this is certainly plausible.

To address the concern that a model with so many parameters might be overfit, I performed dimensionality reduction with FAMD and refit the models using transformed predictor variables. As might be expected in absolute terms the out of sample predictive accuracy (as measured by ELPD) improved, but by an amount much less than the error in the estimation of ELPD. This supports the inferences from the original model. In addition, the composition of the transformed FAMD variables provides some insight into the presentation of sepsis in Blantyre, and their effect on outcome. Though the situation is clearly complex, the composition of the FAMD variables seems to suggest a number of different organ dysfunction syndromes: shock with HIV immunosuppression (FAMD axis one); renal failure with low oxygen saturations (FAMD axis two); low GCS with high white cell count (FAMD axis three); and tachycardia and fever with young age (FAMD axis four). FAMD axes one is strongly associated with death, and FAMD axes two and three more weakly associated. This is in keeping with what is known about sepsis - that sepsis severity, quantified by organ dysfunction, is a major determinant of outcome - but provides a hint that organ dysfunction syndromes may cluster together and, in the case of FAMD 1, that the shock syndrome clusters with HIV status. The dataset is too small for such inferences to be robustly drawn, but this is clinically plausible: the presentation of pneumonia is different from meningitis, for example, in terms of organ dysfunction, and machine learning and clustering techniques applied to large datasets are beginning to identify different sepsis phenotypes[326]. The clinical utility of this remains to be seen.

As described in Chapter 3, time-to-antibacterials is thought to be one of the major modifiable determinants of death in sepsis in high-income settings, but in a crude (unadjusted) analysis there was no signal that this was the case in this cohort. I have expanded that analysis here, and adjusted for putative confounders, but the findings are unchanged. Fitting a logistic regression model and modelling time-to-antibacterials with a linear effect (and adjusting for confounders) found an adjusted odds ratio of 1.01 for death per hour delay with a 95% credible interval (95% CrI 0.98 - 1.04) that includes a clinically relevant effect size (i.e. the aOR of 1.04 [1.02-1.05] from the large observational study of sepsis care in New York[7]). There was a some suggestion from the raw data that a lengthy delay in antibacterials could be associated with an increase in mortality, perhaps with a late nonlinear effect; however the estimates from a nonlinear (second order polynomial) model had such high uncertainty that it is not possible to draw any firm conclusions. The analysis I present



here, therefore, is consistent both with time-to-antibacterials having a similar effect on outcome as in high-income settings, and with no effect. Larger studies would be needed to distinguish these two scenarios. Similarly, uncertainties in estimates of effect of IV fluid administration are such that the estimates could contain a clinically relevant effect, and the analysis presented here does not advance the conclusions drawn in Chapter 3.

The protective effect of TB therapy remains significant and robust to correction for the putative confounders included in the analyses presented here. Using two techniques - logistic regression and propensity score matching - results in the same conclusions: that receipt of TB therapy was associated with survival in this cohort. The effect size of receipt of TB therapy (aOR 0.17 (0.03-0.74)) is greater than the effect of antibacterial delay, even in high income settings, highlighting the fact that determinants of mortality in sepsis are likely to be different in sSA. Subgroup analysis found that the effect is perhaps driven by a mortality benefit in the immunosuppressed and/or anaemic, and the effect size in these subgroups was greater than in those with a diagnosis of TB, albeit with wide confidence intervals. If true, this would suggest that there is a benefit to empiric TB therapy outside those in whom a diagnosis of TB has been made in this study, and could contribute to the equipoise needed to consider clinical trials of empiric TB therapy in sepsis in sSA. This tentative suggestion (numbers are small and confidence intervals wide) is dependent on the diagnostic accuracy of any tests used; the recently developed fujiLAM urinary LAM test shows significantly higher sensitivity than the Alere LAM test used here[327], which could negate the benefit of empiric TB therapy, but would in turn present a powerful point of care triage tool that would permit targeted treatment of disseminated tuberculosis at presentation.

#### 4.5.1 Limitations

I have used a number of techniques to account for confounding, but this will only address the included variables; it is very likely that there are unmeasured confounders, and these could seriously bias the conclusions drawn. This is certainly possible, but any confounder that acted to produce a spurious association between TB therapy and survival would have to be associated with both TB therapy and survival. It seems likely that the clinicians looking after the participants in this study would be more likely to treat sicker patients with TB therapy, and produce bias in the opposite direction. Nevertheless, unmeasured confounders are just that - unmeasured - and it is not possible to address this question with the data here, though this is of course true of any similar study. Given that almost all the participants who received TB therapy were HIV-infected, all conclusions regarding this should be largely applied to people living with HIV; it is not clear (and perhaps unlikely) that the associations described here would also be present in the HIV-uninfected.

I have made an attempt to put the modelling here in the framework of causal inference. If the hypothesised causal framework presented above was right and I had truly adjusted for all confounders, then the estimates of effect I present would be true causal effects. In fact this is unlikely: both because of unmeasured confounders, and that the causal pathways I have hypothesised are almost certainly a oversimplification. I have made choices about inclusion of variables both in the logistic regression and propensity score analysis which could also introduce bias.

## 4.6 Conclusions and further work

In conclusion, the findings from the modelling work in Chapter 3 are largely unchanged when the models are expanded to address possible bias from excluding parameters due to separation, missing data bias, and overfitting. I have identified major associations of mortality in terms of causative pathogen, disease severity, and therapy delivered. Malaria and meningitis are strongly associated with survival to and death by 28 days respectively; BSI seems to be associated with survival to 28 days. Well-recognised markers of disease severity are associated with death: shock, hypoxia, and inability to stand. The FAMD analysis is suggestive that different types of organ dysfunction tend to cluster together. The major association between treatments received and survival is the receipt of TB therapy, in contrast to high income settings where time to antibacterials is the major determinant of survival. The association is stronger in those who are anaemic and immunosuppressed. This is a finding that deserves further exploration; the place of early TB therapy in the treatment of HIV-infected participants with sepsis is unknown, but the data presented here may contribute to the equipoise needed for clinical trials.

## 4.7 Appendix

Below I show bivariable associations of receipt of TB therapy and variable distributions of the propensity-score matched cohort.

Table 4.3: Bivariable associations of receipt of TB treatment in sepsis

Variable	TB treatment	No TB treatment	Difference (95
Host Variables			
Age (years)	37.7 (32.5-42.9)	35.6 (26.8-43.6)	-2.2 (-5.3 to 2.5)
Male sex	30/53 (57%)	84/172 (49%)	-8% (-24 to 9%)
<b>HIV Infected*</b>	<b>46/52 (88%)</b>	<b>97/161 (60%)</b>	<b>-28% (-41 to -15%)</b>
Taking ART <sup>†</sup>	35/46 (76%)	82/97 (85%)	8% (-7 to 24%)
<b>CD4 count<sup>†</sup> (<math>\mu\text{L}^{-1}</math>)</b>	<b>60.0 (26.2-114.8)</b>	<b>123.0 (39.0-274.0)</b>	<b>63.0 (15.0 to 140.5)</b>
<b>Haemoglobin (<math>\times 10^9</math> g dL<sup>-1</sup>)</b>	<b>9.7 (7.4-11.3)</b>	<b>11.1 (8.6-13.9)</b>	<b>1.4 (0.5 to 2.9)</b>
Severity Variables			
Temperature ( $^{\circ}\text{C}$ )	38.5 (38.0-39.2)	38.4 (37.9-39.0)	-0.1 (-0.4 to 0.2)
<b>Heart rate (<math>\text{min}^{-1}</math>)</b>	<b>125.0 (110.0-134.0)</b>	<b>119.5 (99.8-132.0)</b>	<b>-5.5 (-13.0 to -0.5)</b>
Systolic BP (mmHg)	92.0 (81.0-107.0)	99.0 (86.0-120.0)	7.0 (-3.0 to 12.5)
Diastolic BP (mmHg)	67.0 (56.0-71.0)	65.0 (57.0-78.8)	-2.0 (-5.0 to 4.0)
Mean arterial BP (mmHg)	76.0 (65.3-83.7)	77.2 (65.1-91.2)	1.2 (-5.0 to 8.3)
Respiratory rate ( $\text{min}^{-1}$ )	34.0 (30.0-38.0)	34.0 (32.0-37.0)	0.0 (-2.0 to 2.0)
Oxygen saturation (%)	96.0 (94.0-98.0)	96.0 (95.0-98.0)	0.0 (-1.0 to 1.0)
GCS	15.0 (15.0-15.0)	15.0 (15.0-15.0)	0.0 (0.0 to 0.0)
Unable to stand	13/53 (25%)	50/172 (29%)	5% (-10 to 19%)
Lactate (mmol L <sup>-1</sup> )	3.2 (2.4-4.9)	3.4 (2.2-5.3)	0.2 (-0.9 to 0.8)
White cell count ( $\times 10^9$ L <sup>-1</sup> )	6.4 (4.6-9.1)	6.6 (4.3-11.7)	0.2 (-1.3 to 1.8)
Platelet count ( $\times 10^9$ L <sup>-1</sup> )	225.5 (146.8-303.2)	215.0 (145.0-296.0)	-10.5 (-65.0 to 21.0)
<b>Bicarbonate (mmol L<sup>-1</sup>)</b>	<b>18.0 (16.0-21.0)</b>	<b>20.0 (17.0-22.5)</b>	<b>2.0 (0.0 to 3.0)</b>
Urea (mmol L <sup>-1</sup> )	5.0 (3.8-8.7)	4.6 (3.3-7.7)	-0.4 (-2.0 to 0.3)
Creatinine (mmol L <sup>-1</sup> )	76.0 (59.0-105.0)	75.5 (59.0-102.2)	-0.5 (-14.0 to 11.0)
Diagnosis			
BSI	4/53 (8%)	20/172 (12%)	4% (-6 to 14%)
<b>TB</b>	<b>28/53 (53%)</b>	<b>48/172 (28%)</b>	<b>-25% (-41 to -9%)</b>
Malaria	6/53 (11%)	15/172 (9%)	-3% (-13 to 8%)
Meningitis	1/53 (2%)	3/172 (2%)	0% (-4 to 4%)
<b>No diagnosis</b>	<b>15/53 (28%)</b>	<b>96/172 (56%)</b>	<b>28% (12 to 43%)</b>
Treatment Received			
Antibacterials	47/53 (89%)	160/172 (93%)	4% (-6 to 15%)
Time to Antibacterials (hr)	5.1 (3.8-9.7)	5.4 (3.6-13.4)	0.3 (-2.1 to 1.4)
Antifungals	8/53 (15%)	18/172 (10%)	-5% (-17 to 7%)
Time to Antifungals (hr)	45.4 (25.2-60.6)	50.9 (33.2-78.3)	5.5 (-20.3 to 45.9)

Table 4.3: Bivariable associations of receipt of TB treatment in sepsis (*continued*)

Variable	TB treatment	No TB treatment	Difference (95
Antimalarials	6/53 (11%)	6/172 (3%)	-8% (-18 to 2%)
Time to Antimalarials (hr)	4.5 (3.0-11.7)	12.5 (3.3-21.7)	NA
IV fluid (ml)	1.5 (1.0-2.0)	1.2 (0.5-2.0)	-0.3 (-0.8 to 0.1)

*Note:*

BP = Blood pressure, GCS = Glasgow coma scale. Numeric variables are summarised as median (IQR) and categorical variables as proportions. Variable values in the two groups (those who received TB treatment versus not) are compared using difference of medians with 95

\* Participants with HIV status unknown not included in this row

† Includes only HIV-infected participants

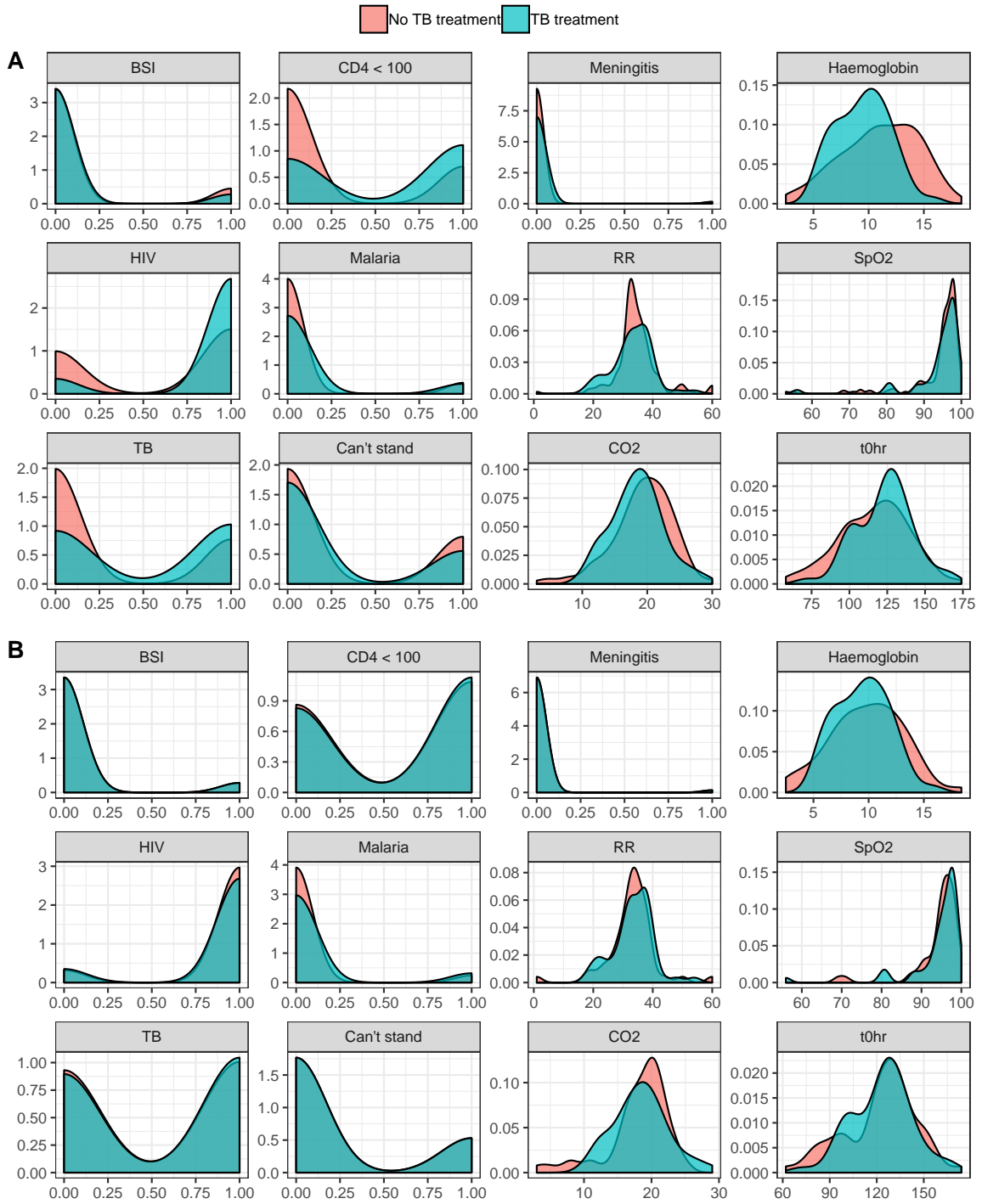


Figure 4.7: Variable distributions following propensity score matching. A: original cohort. B: Propensity score matched cohort. BSI = bloodstream infection, SpO2 = Capillary oxygen saturation, TB = tuberculosis, RR = Respiratory rate, CO2 = Serum bicarbonate, t0hr = Heart rate. Categorical variables (BSI, CD4 < 100, Meningitis, HIV, Malaria, TB, Can't stand) are coded as 1 for present and 0 for absent.



## Chapter 5

# ESBL-E carriage in Malawian adults in health and disease

### 5.1 Chapter Overview

This chapter presents the longitudinal ESBL-E colonisation status of sepsis survivors and two comparator cohorts: antimicrobial-unexposed inpatients and community members. In total, 425 participants were recruited: 225 participants with sepsis, and 100 each of antimicrobial-unexposed inpatients and community members. Stool was sampled at 5 time points over six months. 1416 stool samples were collected and 51% (723/1417) of samples grew 1032 bacteria, most commonly *Escherichia coli* ( $n = 686$ ) and *Klebsiella pneumoniae* ( $n = 245$ ). Baseline ESBL-E carriage prevalence was 49% (95% CI 42-56%) in participants with sepsis, 41% (95% CI 32-52%) and in antimicrobial-unexposed inpatients (both on the day of admission) and 28% (95% CI 20-38%) in “healthy” community members. In multivariable modelling, receipt of cotrimoxazole preventative therapy (CPT), hospitalisation with the previous 4 weeks, use of unprotected water sources, household crowding, and sample collection during the rainy season were all associated with ESBL-E colonisation at enrolment. This suggests that in the community, within-household person to person as well as environmental transmission may be important.

ESBL-E carriage prevalence rose rapidly after admission in antimicrobial-exposed participants, to 78% (95% CI 71-84%) by the day 7 visit, a rise which was not seen in the antimicrobial-unexposed arm of the study (51% at day 7 visit, 95% CI 38-64%), suggesting that antimicrobial exposure and not hospitalisation per se is driving carriage. However, this conclusion is open to confounding because antimicrobial-exposed participants differ from antimicrobial-unexposed in a number of important ways, including a lower HIV prevalence and shorter median length of hospital stay. Attempts to control for this confounding

with logistic regression failed: collinearity and a small dataset resulted in very uncertain parameter estimates. In addition, aggregate prevalences obscure a complex pattern of shifting between colonised and non-colonised states at the individual level, and it is not clear whether this represents intermittent ESBL-E shedding, imperfectly sensitive tests, or true acquisition and loss events on a short time scale. I describe these difficulties, and outline the methods that will be used to address them over the next three chapters.

## 5.2 Introduction and chapter aims

In the preceding two chapters I have described significant exposure to broad-spectrum antimicrobials in a sepsis cohort. I now turn my attention to the consequences of this antimicrobial exposure in terms of acquisition and carriage of ESBL-E and present a description of longitudinal carriage of ESBL-E in sepsis survivors, plus two comparator cohorts of antimicrobial unexposed inpatients and community members. I also attempt to describe risk factors for carriage.

Data from sSA suggest that ESBL-E gut mucosal carriage is common (Chapter 1), but routes of transmission are unknown. Antimicrobial exposure and hospitalisation have been associated with ESBL-E carriage, but such longitudinal cohort data that do exist are associated with short follow up time, often not following participants beyond hospital discharge. The role of hospital acquisition of ESBL-E and carriage into the community in driving drug-resistant infection versus community transmission is therefore unclear. In addition, the mechanism of ESBL-E acquisition in hospitalised adults in sSA including the relative effects of antimicrobials versus hospitalisation - exactly the understanding that would be needed to design effective interventions - is unknown.

The aims of this chapter are therefore threefold:

1. To present the details of recruitment, follow up and ESBL-E colonisation status of the participants recruited to the clinical study underpinning this thesis;
2. To explore associations of baseline ESBL-E colonisation to understand potential community ESBL-E transmission routes;
3. To explore associations of ESBL-E acquisition by 28 days, particularly the relative effects of hospitalisation and antimicrobial exposure.

## 5.3 Methods

The methods for recruitment and follow up of the clinical cohort along with the laboratory methods of sample processing, stool culture, bacterial identification, ESBL confirmation



and antimicrobial sensitivity testing are described in Chapter 2, Methods. Further methods of the statistical analysis carried out in this chapter are detailed here.

Summary statistics across the three arms of the study are presented as proportions for categorical data and medians and interquartile ranges for continuous data, with p-values from Fisher's exact test and the Kruskal-Wallis test, respectively, used to test for differences between the arms. The magnitude of the time-varying exposures of interest - hospitalisation and antimicrobial exposure - were expressed in three ways, to ensure that the study procedures had generated three arms with good separation of exposures:

1. The proportion of participants in each arm who were exposed to a given exposure on any given day was plotted;
2. Total person-days of exposure for each arm was calculated; the person-days at risk were not equal across the three arms of the study because of varying numbers of participants and drop-out rates, so the total person-days at risk were also calculated;
3. The number of participants who were exposed to a given exposure, along with the median length of exposure were calculated.

ESBL-E colonisation was expressed as a simple proportion at each time point, and visualised by plotting the proportion with binomial confidence intervals. For these plots, Arm 2 and 3 participants were censored at first antimicrobial exposure (Arm 2 and 3) and hospitalisation (Arm 3). Whilst time of measurement of ESBL-E status was ostensibly at day 0,7,28,90 or 180, in reality it was distributed around these points. To visualise these data accounting for this, ESBL-E carriage status was plotted against time with ESBL-E colonisation coded as 1 and not colonised as 0, and a non-parametric LOESS regression line (with first order polynomial and smoothing parameter 0.75) fitted to them. This fits a local smoothed linear regression using least squares and a proportion of the of the data points, to produce a smoothed local best fit curve through the data points; in this case broadly equivalent to binning observations in a large number of time category bins and calculating a prevalence for each one, and can be interpreted as a ESBL-E rolling prevalence which accounts for the varying measurement time.

This does not, however, account for the fact that the measurements are clustered within individuals. In order to do this, a two-state model (with ESBL-E colonised and uncolonised states) was fitted using the Aalen-Johansen estimate of state occupancy in the *survival* package in R. This is a generalisation on the Kaplan-Meier curve and allows plotting of state occupancy probability - an estimate of carriage prevalence at any time point - as a function of time, rather than just the survival function as in a Kaplan-Meier curve. However, the Aalen-Johansen estimate assumes that the time of state transition is known. In fact, the data considered here are interval censored - that is, that transition is only known to have happened within a particular time period between two study visits, and so to generate

this estimate of state occupancy it is necessary to assume a transition time: I assumed it happened halfway between measurements. A Markov model can account for all these difficulties; the development and fitting of such a model is the subject of Chapter 8. ESBL-E carriage was also visualised as a heat map, with each cell representing a sample, columns representing individuals and rows the study visits.

To explore the associations of ESBL-E carriage, two logistic regression models were fit. The first aimed to explore associations of baseline ESBL-E carriage. Variables that I hypothesised *a priori* would be related to ESBL-E carriage were included in the model: age, sex, HIV status, study arm, receipt of antiretroviral therapy (ART) or cotrimoxazole preventative therapy (CPT), hospitalisation or receipt of antimicrobial therapy, household crowding (number of adults and number of children in the household separately as linear continuous variables), presence of animals at home, presence of a flushing toilet at home, use of unprotected water sources (defined as surface water or unprotected springs or wells), whether water was treated with chlorine, and sample collection during rainy season (defined as date of sample collection between 1st November and 30th April). These variables were fit individually (univariable model) and then in a full multivariable model. Odds ratios with 95% confidence intervals and p-values are presented.

The second logistic regression model aimed to explore associations of ESBL-E acquisition by the 28 day visit. This analysis included only participants who were a) ESBL-E uncolonised at baseline and b) had an available sample at 28 days  $\pm$  2 weeks. To explore associations of ESBL-E acquisition, the exposures of interest (again specified *a priori*: ceftriaxone, amoxicillin, ciprofloxacin and cotrimoxazole exposure and hospitalisation) were first quantified as days of exposure between the baseline and follow up visit, and binned into three groups: no exposure, five or fewer days of exposure, or more than five days of exposure. Proportion of participants who had a detectable ESBL-E at follow up in each group were plotted, stratified by these groups, to assess both for any association and for a dose-response relationship. Multivariable analysis was carried out by dichotomising each exposure as a binary variable (exposed/non-exposed) and including them all in a logistic regression model. They were not used as a continuous linear predictor because a linear relationship between antimicrobial exposure and ESBL-E acquisition seems unlikely, and there were not enough data points to categorise the variables with more granularity.

## 5.4 Results

### 5.4.1 Study population

In total, 425 participants were recruited to the study between 19th February 2017 and 2nd October 2018; 225 participants with sepsis (Arm 1), 100 inpatients without antimicrobial

exposure at baseline (Arm 2) and 100 community members (Arm 3). Flow of participants through the study is shown in Figure 5.1. It was often challenging to collect stool samples from participants but 87% (1416/1631) eligible patient-visits resulted in the collection of a stool sample. Drop out from the study and failure to collect stool samples were similar in arm 1 and 2 and with no apparent systematic bias, but both drop out and missing samples were less frequent in arm 3 (Figure 5.2A). There was significant variation in the timing of stool sample collection, with a broad distribution around the ostensible collation day (Figure 5.2B).

The baseline characteristics of the enrolled participants are shown in Table 5.1. There were some important differences between the arms of the study: despite matching on age and sex, antimicrobial-unexposed participants were older. They were also less likely to be HIV-infected than participants with sepsis (13% [12/89] of those with known HIV status were HIV-infected versus 67% [143/213] with sepsis), and less likely to have been treated for TB. Sepsis participants were more likely to have received antimicrobials or been hospitalised in the previous 4 weeks. In the community arm of the study, there were a high proportion of participants (60% [60/100]) with an unknown HIV status, and there were some differences in toilet facilities, water sources, cooking fuel and presence of animals at home across the three groups.

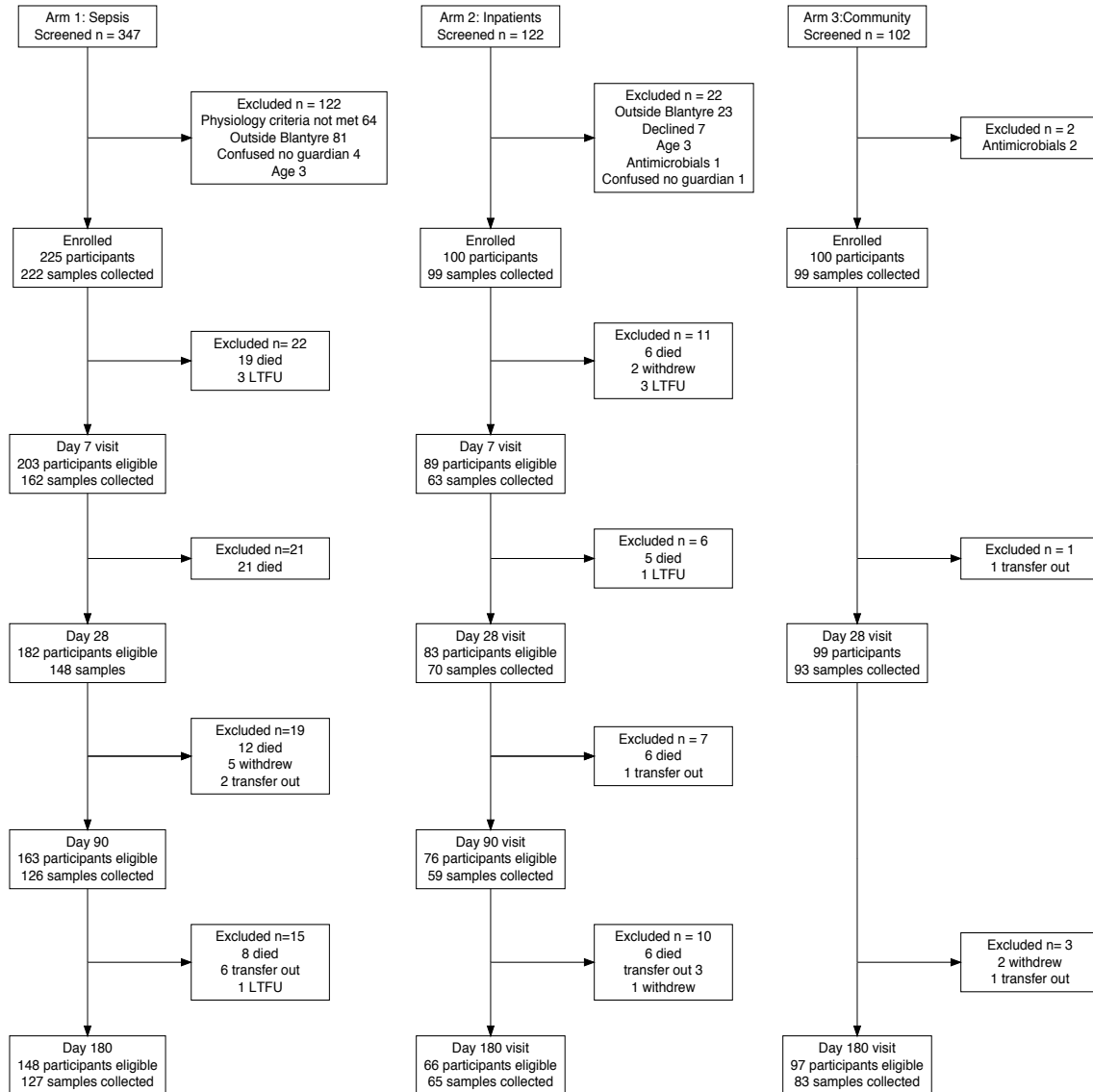


Figure 5.1: Study recruitment and follow up. At each time point *eligible participants* refers to participants who are known to be alive and have not withdrawn from the study by that time point, and *samples collected* refers to patients from whom a stool sample was successfully collected for that visit.

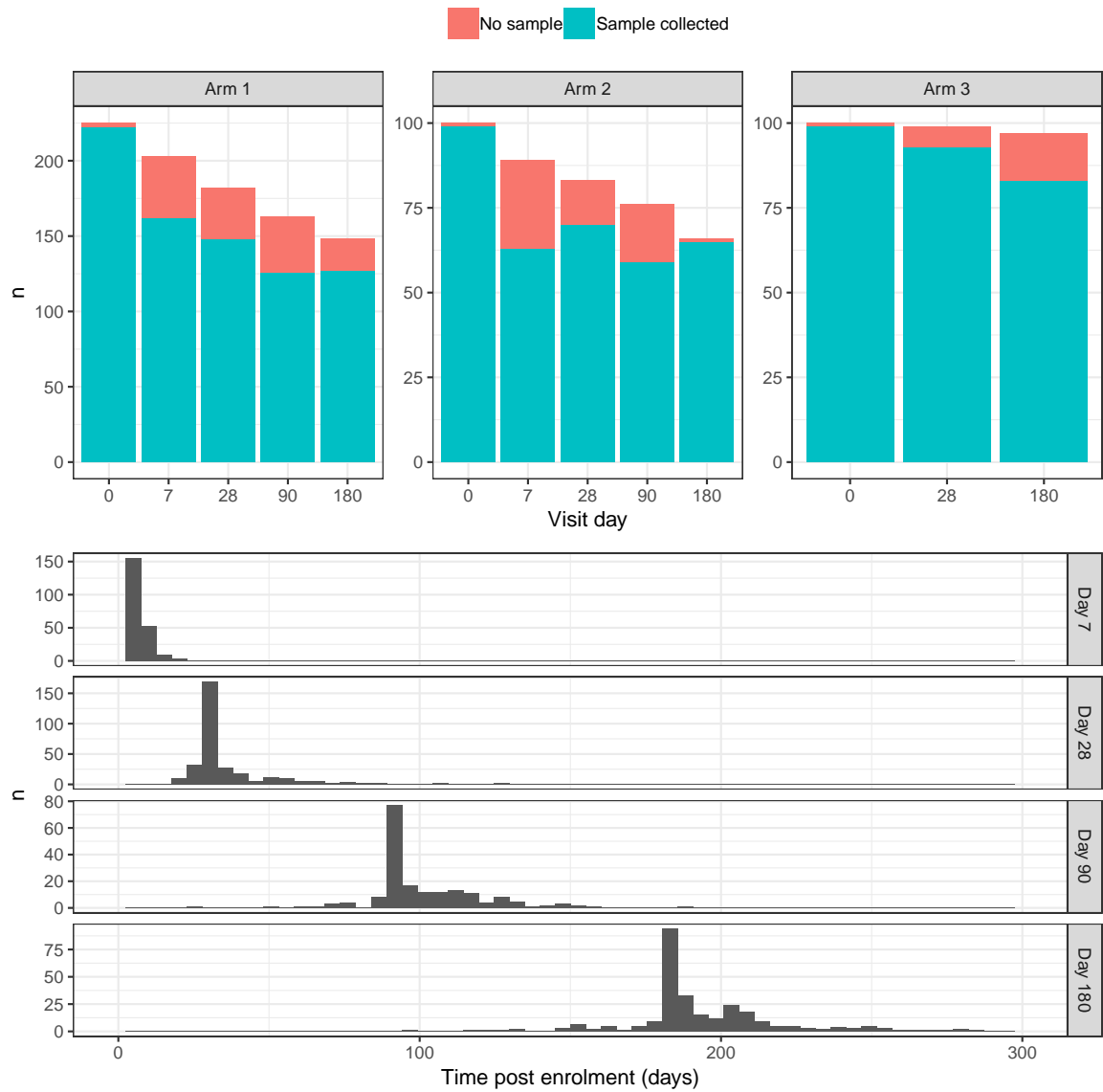


Figure 5.2: A: Missing stool samples stratified by arm and visit. Bar height at a given visit represents the number of eligible participants, coloured by successful sample collection (blue) or failure to collect a sample (red). B: Distribution of actual day of sample collection for ostensible day 7, 28, 90 and 180 samples showing considerable variation.

Table 5.1: Participant characteristics

Variable	Sepsis	Inpatient	Community	p	Total
<b>Demographics</b>					
Age (yr)	<b>35.9 (27.8-43.5)</b>	<b>40.4 (29.1-48.3)</b>	<b>32.5 (24.0-38.4)</b>	<b>&lt;0.001</b>	<b>35.6 (26.9-43.9)</b>
Male sex	114/225 (51%)	51/100 (51%)	40/100 (40%)	0.533	205/425 (48%)
<b>HIV/TB status</b>					
HIV Reactive	<b>143/225 (64%)</b>	<b>12/100 (12%)</b>	<b>18/100 (18%)</b>	<b>&lt;0.001</b>	<b>173/425 (41%)</b>
HIV Non Reactive	<b>70/225 (31%)</b>	<b>77/100 (77%)</b>	<b>22/100 (22%)</b>	<b>&lt;0.001</b>	<b>169/425 (40%)</b>
HIV Unknown	<b>12/225 (5%)</b>	<b>11/100 (11%)</b>	<b>60/100 (60%)</b>	<b>&lt;0.001</b>	<b>83/425 (20%)</b>
Ever treated for TB	<b>37/225 (16%)</b>	<b>5/100 (5%)</b>	<b>4/100 (4%)</b>	<b>0.002</b>	<b>46/425 (11%)</b>
Of those, current TB treatment	10/37 (27%)	0/5 (0%)	4/4 (100%)	0.098	14/46 (30%)
<b>ART status*</b>					
Current ART*	117/143 (82%)	9/12 (75%)	18/18 (100%)	0.859	144/173 (83%)
Months on ART	28.7 (3.7-72.6)	35.1 (2.9-79.8)	31.5 (13.0-79.9)	0.693	29.5 (3.8-72.8)
ART regimen: EFV/3TC/TDF	110/117 (94%)	8/9 (89%)	17/18 (94%)	1.000	135/144 (94%)
<b>ART status</b>					
Current CPT <sup>†</sup>	98/141 (70%)	5/12 (42%)	7/18 (39%)	0.328	110/171 (64%)
<b>Healthcare exposure last 4wk</b>					
Antibiotics	<b>60/225 (27%)</b>	<b>0/100 (0%)</b>	<b>0/100 (0%)</b>	<b>&lt;0.001</b>	<b>60/425 (14%)</b>
Hospitalised	<b>18/225 (8%)</b>	<b>1/100 (1%)</b>	<b>0/100 (0%)</b>	<b>0.001</b>	<b>19/425 (4%)</b>
<b>Tobacco/alcohol use</b>					
Never tobacco	196/225 (87%)	93/100 (93%)	90/100 (90%)	0.929	379/425 (89%)
Ex tobacco	17/225 (8%)	6/100 (6%)	2/100 (2%)	0.180	25/425 (6%)
Current tobacco	12/225 (5%)	1/100 (1%)	8/100 (8%)	0.070	21/425 (5%)
Current alcohol	51/225 (23%)	16/100 (16%)	18/100 (18%)	0.502	85/425 (20%)
<b>Education</b>					
Primary incomplete or complete	97/225 (43%)	50/100 (50%)	42/100 (42%)	0.739	189/425 (44%)
Some secondary education	47/225 (21%)	18/100 (18%)	30/100 (30%)	0.238	95/425 (22%)
Secondary school complete	48/225 (21%)	16/100 (16%)	19/100 (19%)	0.677	83/425 (20%)
No formal schooling	16/225 (7%)	13/100 (13%)	4/100 (4%)	0.094	33/425 (8%)
College or higher	17/225 (8%)	3/100 (3%)	5/100 (5%)	0.346	25/425 (6%)
<b>Employment</b>					
Unemployed	82/225 (36%)	34/100 (34%)	32/100 (32%)	0.866	148/425 (35%)
Self-employed	56/225 (25%)	32/100 (32%)	35/100 (35%)	0.325	123/425 (29%)
Currently employed	65/225 (29%)	26/100 (26%)	18/100 (18%)	0.269	109/425 (26%)
Student	21/225 (9%)	6/100 (6%)	15/100 (15%)	0.153	42/425 (10%)

Retired	1/225 (0%)	2/100 (2%)	0/100 (0%)	0.280	3/425 (1%)
<b>Toilet facilities</b>					
Pit latrine with slab +/- foot rest	104/225 (46%)	25/100 (25%)	35/100 (35%)	0.039	164/425 (39%)
Pit latrine with slab and cover +/- foot rest	45/225 (20%)	19/100 (19%)	55/100 (55%)	<0.001	119/425 (28%)
Hanging toilet/latrine	59/225 (26%)	48/100 (48%)	9/100 (9%)	<0.001	116/425 (27%)
Flush Toilet (any type)	14/225 (6%)	5/100 (5%)	1/100 (1%)	0.118	20/425 (5%)
No toilet	2/225 (1%)	2/100 (2%)	0/100 (0%)	0.533	4/425 (1%)
Composting toilet	1/225 (0%)	1/100 (1%)	0/100 (0%)	0.720	2/425 (0%)
<b>Main water source</b>					
Public tap/standpipe	51/225 (23%)	8/100 (8%)	66/100 (66%)	<0.001	125/425 (29%)
Piped outside dwelling	69/225 (31%)	37/100 (37%)	9/100 (9%)	<0.001	115/425 (27%)
Tube well/borehole	64/225 (28%)	35/100 (35%)	15/100 (15%)	0.032	114/425 (27%)
Piped into dwelling	30/225 (13%)	11/100 (11%)	7/100 (7%)	0.353	48/425 (11%)
Unprotected well/spring	5/225 (2%)	6/100 (6%)	2/100 (2%)	0.181	13/425 (3%)
Surface water (including rainwater collection)	4/225 (2%)	2/100 (2%)	0/100 (0%)	0.556	6/425 (1%)
Tube well with powered pump	2/225 (1%)	1/100 (1%)	1/100 (1%)	1.000	4/425 (1%)
Treat water with chlorine	19/225 (8%)	5/100 (5%)	0/100 (0%)	0.004	24/425 (6%)
<b>No. household members</b>					
Children	2.0 (1.0-3.0)	2.0 (1.0-3.0)	2.0 (1.0-3.0)	0.395	2.0 (1.0-3.0)
Adults	2.0 (2.0-3.0)	3.0 (2.0-4.0)	2.0 (2.0-4.0)	0.907	3.0 (2.0-4.0)
<b>Electricity</b>					
Electricity available in house	119/225 (53%)	41/100 (41%)	58/100 (58%)	0.357	218/425 (51%)
<b>Main cooking fuel</b>					
Charcoal	161/225 (72%)	63/100 (63%)	88/100 (88%)	0.291	312/425 (73%)
Wood	61/225 (27%)	35/100 (35%)	11/100 (11%)	0.004	107/425 (25%)
Electricity	3/225 (1%)	2/100 (2%)	1/100 (1%)	0.869	6/425 (1%)
<b>Animals at home?</b>					
Any animal	71/225 (32%)	43/100 (43%)	15/100 (15%)	0.004	129/425 (30%)
Poultry	46/71 (65%)	34/43 (79%)	10/15 (67%)	0.800	90/129 (70%)
Dogs	18/71 (25%)	11/43 (26%)	9/15 (60%)	0.201	38/129 (29%)
Other	11/71 (15%)	9/43 (21%)	5/15 (33%)	0.413	25/129 (19%)
Goats	12/71 (17%)	7/43 (16%)	1/15 (7%)	0.830	20/129 (16%)
Cattle	2/71 (3%)	3/43 (7%)	0/15 (0%)	0.406	5/129 (4%)

*Note:*

ART = Antiretroviral therapy, CPT = Cotrimoxazole preventative therapy, EFV: Efavirenz, 3TC: Lamivudine, TDF: Tenofovir. Numeric values are median (IQR)) unless otherwise stated. P-values are to assess for differences across the three groups: Fisher's exact test across the groups for categorical variable, and Kruskal-Wallace test for continuous variables.

\* ART status includes HIV reactive only as denominator

† Missing CPT data for two participants.



### 5.4.2 Exposures during the study period

Exposures to antimicrobials and hospitalisation of the cohort are shown in Figure 5.3 and Table 5.2. Antimicrobial-unexposed inpatients (arm 2 participants) had a shorter length of hospital stay than participants with sepsis (arm 1 participants): median (IQR) 2 (2-7) versus 5 (2-10) days,  $p = 0.002$  by Kruskal-Wallis test. Five of the 100 arm 2 participants were taking CPT at baseline, 18 received further courses of antimicrobials during the study period, and two were started on TB therapy. Some participants received combinations of these therapies, so in total 23% (23/100) arm 2 participants received an antibacterial during the study period, mostly within 30 days following enrolment (Figure 5.3), and most commonly ceftriaxone (Table 5.2). All of these participants were censored on first exposure.

Both antimicrobial exposure and hospitalisation were unusual in the community cohort; 7% (7/100) community (arm 3) participants were taking CPT and one received a 5-day course of amoxicillin meaning that 8% (8/100) Arm 3 participants received an antibacterial during the study period. In addition one arm 3 participant was hospitalised for 1 day in the study period. No arm 3 participant received any TB therapy, and no Arm 2 or 3 participants received any antimalarial or antifungal therapy during the study period.

Because of the chronic nature of the therapy, the greatest antimicrobial exposure (in terms of participant-days) in all arms were to cotrimoxazole and TB therapy, by an order of magnitude (Table 5.2). Apart from these, the most commonly received antibacterial by arm 1 participants was ceftriaxone by some distance with 998 participant-days of exposure in 189 participants during the study period, and a median 5 (IQR 3-7) day course. Ciprofloxacin and amoxicillin were also commonly received, with 61 participants receiving 398 participant-days of exposure to ciprofloxacin with a median 7 (IQR 5-7) day course, and 39 participants receiving 235 participant-days of exposure to amoxicillin with a median 5 (IQR 5-7) day course.

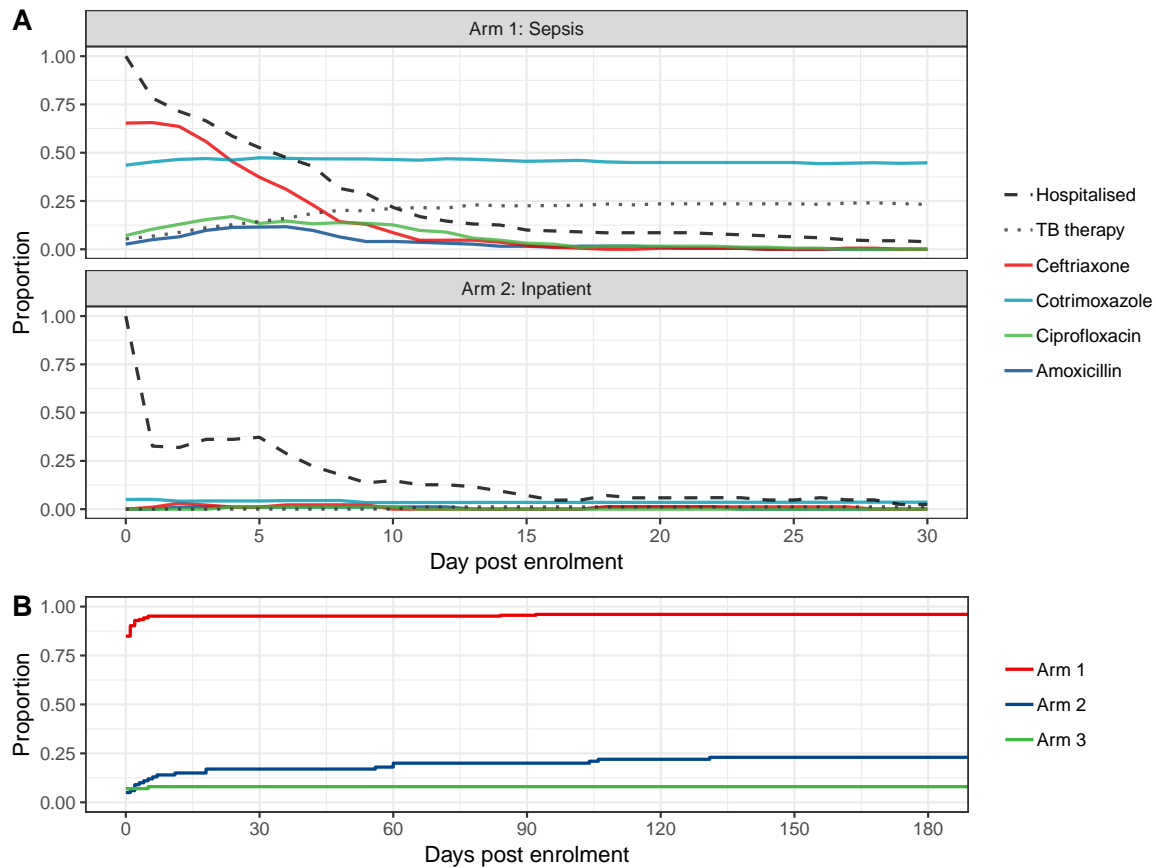


Figure 5.3: Hospital and antibacterial exposure of participants expressed as (A) proportion of arm 1 and arm 2 participant who are hospitalised and/or exposed to the most commonly received antibacterials on any given day and (B) cumulative proportion of participants who have been exposed to any antibacterial over the study period.

Table 5.2: Antimicrobial and hospital exposure stratified by arm

Exposure	Number exposed			Exposure (person-days)			Median (IQR) exposure length (days)		
	Arm 1	Arm 2	Arm 3	Arm 1	Arm 2	Arm 3	Arm 1	Arm 2	Arm 3
Total At Risk	225	100	100	33797	14336	21983	-	-	-
<b>Exposures</b>									
Hospitalised	225	100	1	1727	500	1	5 (2-10)	2 (2-7)	1 (1-1)
Cotrimoxazole	110	6	7	14447	549	1388	180 (27-190)	86 (6-177)	190 (183-206)
TB therapy	52	2	0	6843	291	0	178 (58-180)	146 (133-158)	-
Ceftriaxone	183	7	0	997	26	0	5 (3-7)	3 (2-4)	-
Ciprofloxacin	61	2	0	398	12	0	7 (5-7)	6 (6-6)	-
Amoxicillin	38	3	1	235	21	5	7 (5-7)	5 (5-8)	5 (5-5)
Metronidazole	24	2	0	148	10	0	6 (2-7)	5 (5-5)	-
Fluconazole	27	0	0	118	0	0	3 (2-5)	-	-
Aciclovir	2	0	0	47	0	0	24 (16-31)	-	-
Co-amoxiclav	10	2	0	40	12	0	5 (2-5)	6 (6-6)	-
Erythromycin	5	0	0	38	0	0	7 (5-11)	-	-
Doxycycline	7	0	0	34	0	0	3 (2-6)	-	-
Artesunate	11	0	0	25	0	0	2 (2-3)	-	-
LA	7	0	0	19	0	0	3 (2-3)	-	-
Streptomycin	2	0	0	16	0	0	8 (7-9)	-	-
Gentamicin	4	0	0	15	0	0	4 (3-5)	-	-
Amphotericin	2	0	0	8	0	0	4 (4-4)	-	-
Azithromycin	2	2	0	7	12	0	4 (3-4)	6 (6-6)	-
Penicillin	2	0	0	5	0	0	2 (2-3)	-	-
Flucloxacillin	2	0	0	5	0	0	2 (2-3)	-	-
Chloramphenicol	1	0	0	1	0	0	1 (1-1)	-	-
Quinine	1	0	0	1	0	0	1 (1-1)	-	-

*Note:*

TB = tuberculosis, LA =lumefantrine artemether. Median exposure length includes only those exposed. Total at risk shows the total number of participants and participant-days of follow up included in the study.

Table 5.3: ESBL carriage stratified by arm and visit

Visit	Arm 1 (Sepsis)		Arm 2 (Inpatient)		Arm 3 (Community)	
	n	Any ESBL	n	Any ESBL	n	Any ESBL
Day 0	222	109 (49%)	99	41 (41%)	99	28 (28%)
Day 7	162	127 (78%)	63	32 (51%)	-	-
Day 28	148	106 (72%)	71	37 (52%)	92	29 (32%)
Day 90	126	71 (56%)	60	29 (48%)	-	-
Day 180	127	61 (48%)	65	29 (45%)	83	24 (29%)

### 5.4.3 ESBL-E colonisation

ESBL-E colonisation prevalence as a function of time across the three arms of the study is shown in Table 5.3 and Figure 5.4. Baseline colonisation prevalence was high in all groups, and higher in arm 1 and 2 participants than community members: 49% (95% CI 42-56%) in arm 1 participants, 41% (95% CI 32-52%) in arm 2 and 28% (95% 20-38%) in arm 3. Both hospitalised groups showed a rise in colonisation prevalence following admission, though this is much more marked in arm 1 participants: by the day 7 visit 78% (95% CI 71-84%) of arm 1 participants were colonised compared to 51% (38-64%) of arm 2 participants. This difference persisted through to day 28, when the crude prevalence in arm 1 was 72% (95% CI 64-79%) versus 52% (95% CI 40-64%) in arm 2. By the end of the study period the prevalence had fallen back to baseline levels in both groups. Within an individual, there was often frequent flipping between the ESBL colonised to uncolonised state and back again at different study visits and often on short time-scales (Figure 5.5).

In total, 723/1417 (51%) of samples grew at least one ESBL-E; 1032 organisms were grown from the 723 samples, with a median 1 (IQR [1-2]) ESBL-E per sample. The most commonly isolated organism as identified by the API system was *E. coli* ( $n = 686$ ) followed by *Klebsiella pneumoniae* ( $n = 245$ , Figure 5.6). Antimicrobial sensitivity testing was carried out on the first 694/1032 (67%) organisms. Cotrimoxazole resistance was near universal (675/694 [97%] of isolates), with intermediate proportions of gentamicin (367/694 [53%]) and ciprofloxacin (457/694 [65%]) resistance and less chloramphenicol resistance (232/694 [33%] of isolates). Meropenem and amikacin resistance was very unusual (14/694 [2%] and 15/694 [2%] of isolates respectively) but the fact that resistance to these antibacterials of last resort was seen at all is troubling.

### 5.4.4 Associations of ESBL colonisation

I then used logistic regression to explore associations of ESBL-E colonisation at baseline. Of the 420 participants with an available enrolment stool culture result, 42% (178/420) cultured at least one ESBL-E. Univariable and multivariable associations of colonisation at enrolment are shown in 5.4. In univariable analysis HIV infection, ART and CPT are

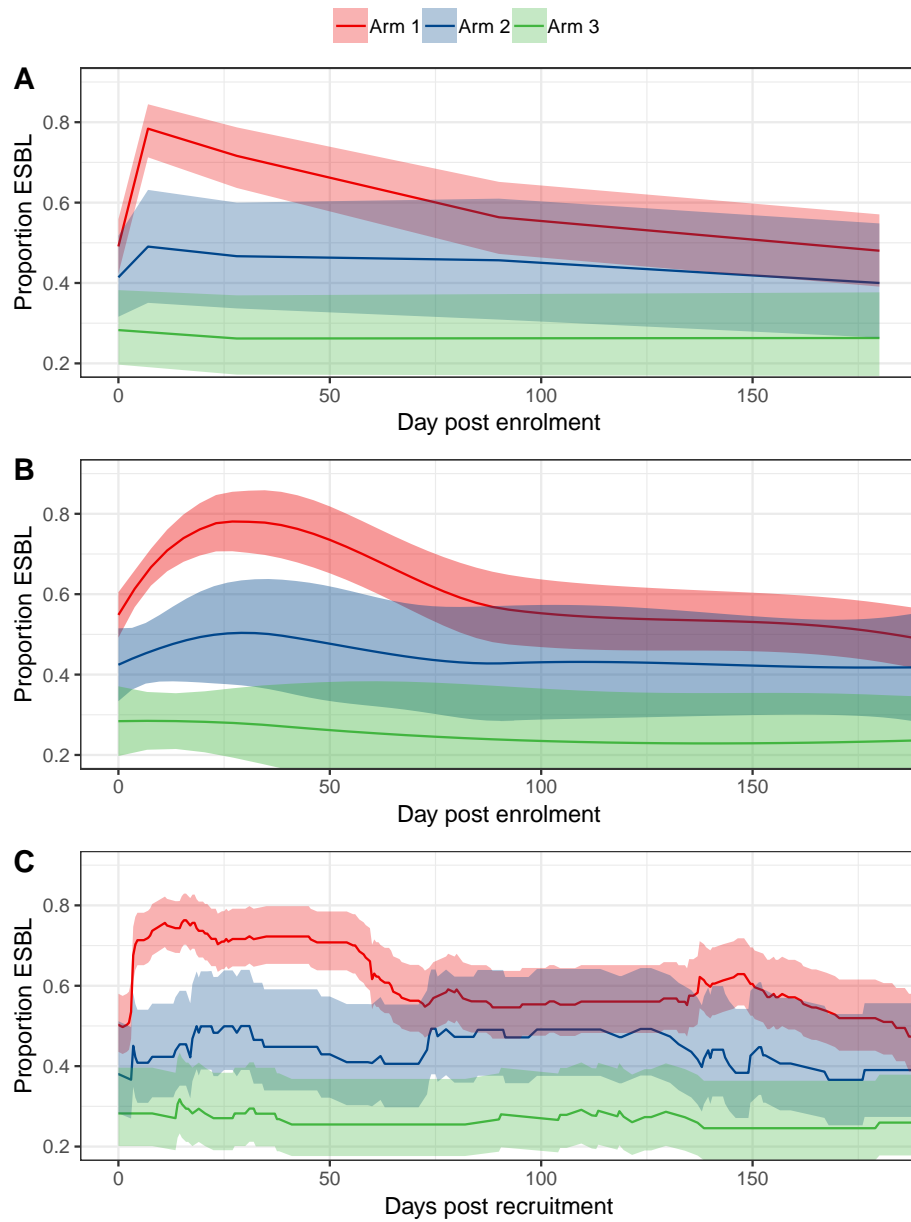


Figure 5.4: ESBL carriage prevalence as a function of time visualised in three different ways. In each case participants from arm 2 are censored on first antimicrobial exposure and arm 3 are censored on first antimicrobial exposure or hospitalisation. Top (A) prevalence at each visit plotted at ostensible visit time; however, the visits are in fact distributed in time themselves so the middle plot (B) is an attempt to show this by fitting a nonparametric smoothed LOESS regression line with a local linear regression. However the confidence intervals in this method are too narrow because they assume independence of the measurements, which are in fact clustered within patients. The bottom panel (C) is an estimate of the proportion of ESBL-colonised participants from the Aalen-Johansen estimate, which is a generalisation of the Kaplan-Meier curve. This takes into account the nonindependence of the measurements, but does not take into account the interval-censored nature of the data, and transitions to and from the ESBL colonised state are hence assumed to happen halfway between measurements.



Figure 5.5: Heatmap showing distribution of stool samples with detectable ESBL-E. Each column represents a patient, and each cell a stool sample with valid result, coloured by presence or absence of detectable ESBL-E, to demonstrate the complex patterns of apparent acquisition and loss in many individuals over the study period.

associated with ESBL-E colonisation. This seems to be largely mediated by CPT as the HIV and ART associations largely disappear on multivariable modelling but the effect of CPT is still apparent (aOR 2.3 [95% CI 1.0 - 5.5]). Hospitalisation within the 4 weeks prior to admission was strongly associated with ESBL-E colonisation on multivariable modelling, though with wide confidence intervals (aOR 5.9 [95% CI 1.8-27.0]). Antimicrobial exposure was not, but with confidence intervals that contained a clinically relevant effect size (aOR 1.3 [95% 0.7 - 2.6]). ESBL-E colonisation was more likely with more adults in the household (aOR 1.2 [95% CI 1.0-1.4] per extra adult), with use of an unprotected water source (aOR 3.0 [95% CI 1.1 - 8.8]) and in the rainy season (aOR 2.2 [95% CI 1.4-3.4]).

To explore associations of acquisition of ESBL-E by the day 28 visit, I analysed only those participants who had no detectable ESBL-E at baseline, and an available follow up samples at 28 days  $\pm$  14 days. These numbered 150 participants: 64 Arm 1, 37 Arm 2 and 49 Arm 3 participants, and 49% (73/150) of them had a detectable ESBL-E at day 28. Bivariable associations of ESBL-E acquisition with antimicrobial and hospital exposures are shown in Figure 5.7A, stratified by the length of exposure; all antibacterials (including TB therapy) showed an association with ESBL-E acquisition, with a suggestion of a dose-response effect (in that longer exposures generally had a higher colonisation prevalence), but confidence intervals were large in many cases. Antimalarials did not show this effect though here uncertainty in the estimates precludes drawing any firm conclusions, as it does

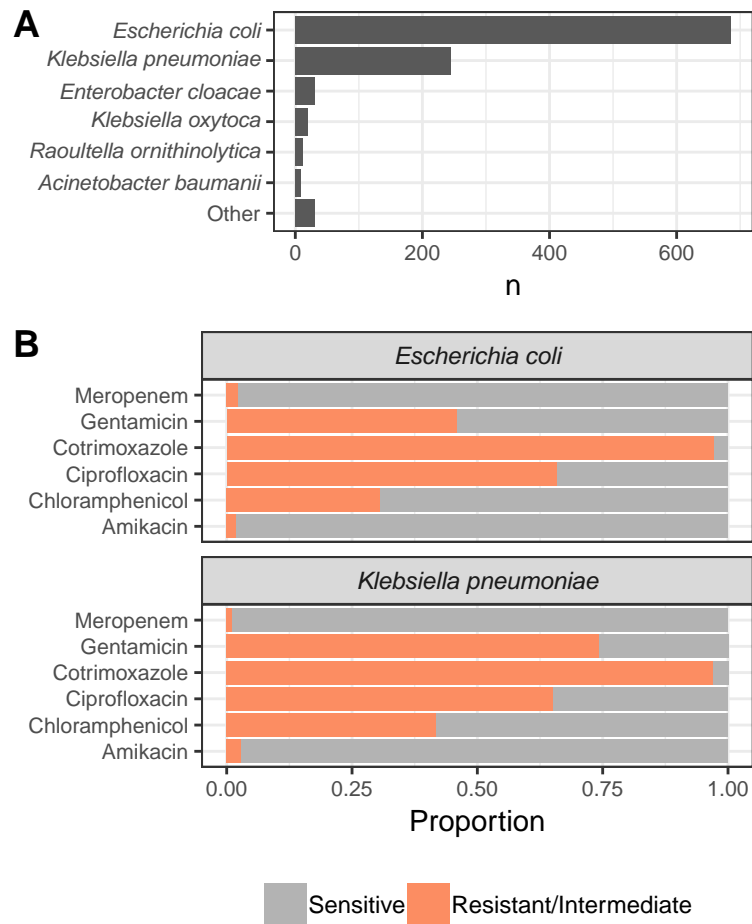


Figure 5.6: Species (A) and antimicrobial sensitivities (B) of cultured ESBL-E

for antifungals. These relationships are very likely confounded, so should be regarded with caution; however, due to a small dataset size and collinearity, logistic regression modelling of ESBL-E acquisition (Figure 5.7B) produces such uncertain parameter estimates that no conclusions can be drawn. A better modelling strategy using continuous time Markov models is presented in Chapter 8.



Table 5.4: Univariable and multivariable associations of ESBL colonisation at enrolment

Variable	Univariable		Multivariable	
	OR (95% CI)	p-value	aOR (95% CI)	p-value
Demographics				
Age (per year)	1.00 (0.99-1.02)	0.709	1.00 (0.98-1.02)	0.898
Male sex (vs female)	1.23 (0.84-1.82)	0.287	1.42 (0.93-2.19)	0.106
Study Arm				
Arm 2 (vs 1)	0.73 (0.45-1.18)	0.203	1.57 (0.84-2.96)	0.157
Arm 3 (vs 1)	<b>0.41 (0.24-0.68)</b>	<b>0.001</b>	0.91 (0.45-1.84)	0.801
HIV status				
HIV+ (vs HIV-)	<b>1.68 (1.09-2.59)</b>	<b>0.018</b>	1.16 (0.46-2.84)	0.750
HIV unknown (vs HIV-)	0.71 (0.40-1.24)	0.229	1.09 (0.55-2.18)	0.798
CPT (vs none)	<b>2.46 (1.58-3.86)</b>	<b>&lt;0.001</b>	2.29 (0.98-5.54)	0.060
ART (vs none)	<b>1.99 (1.32-3.00)</b>	<b>0.001</b>	1.06 (0.35-3.17)	0.918
Exposures last month				
Hospitalisation	<b>7.87 (2.57-34.22)</b>	<b>0.001</b>	<b>5.90 (1.78-26.94)</b>	<b>0.008</b>
Antibiotics*	<b>2.14 (1.27-3.67)</b>	<b>0.005</b>	1.34 (0.71-2.57)	0.368
Household size				
Children (per 1)	1.00 (0.87-1.14)	0.979	0.98 (0.84-1.14)	0.793
Adults (per 1)	1.14 (0.99-1.31)	0.064	<b>1.19 (1.02-1.40)</b>	<b>0.026</b>
Keep animals (vs not)	1.33 (0.88-2.03)	0.176	1.16 (0.73-1.85)	0.527
WaSH behaviour				
Flushing toilet (vs not)	1.38 (0.55-3.44)	0.481	0.94 (0.34-2.55)	0.908
Unprotected water source	2.43 (0.96-6.64)	0.068	<b>2.98 (1.08-8.78)</b>	<b>0.039</b>
Treat water (vs not)	1.16 (0.50-2.66)	0.725	0.94 (0.37-2.34)	0.900
Season				
Rainy season (vs. dry)	<b>2.05 (1.38-3.06)</b>	<b>&lt;0.001</b>	<b>2.17 (1.38-3.44)</b>	<b>0.001</b>

*Note:*

CPT = Cotrimoxazole preventative therapy, ART = antiretroviral therapy, WaSH = Water, sanitation and hygiene. Entries in bold are those for which 95% confidence intervals do not cross 1.

\* Antibiotics includes TB therapy but excludes CPT.

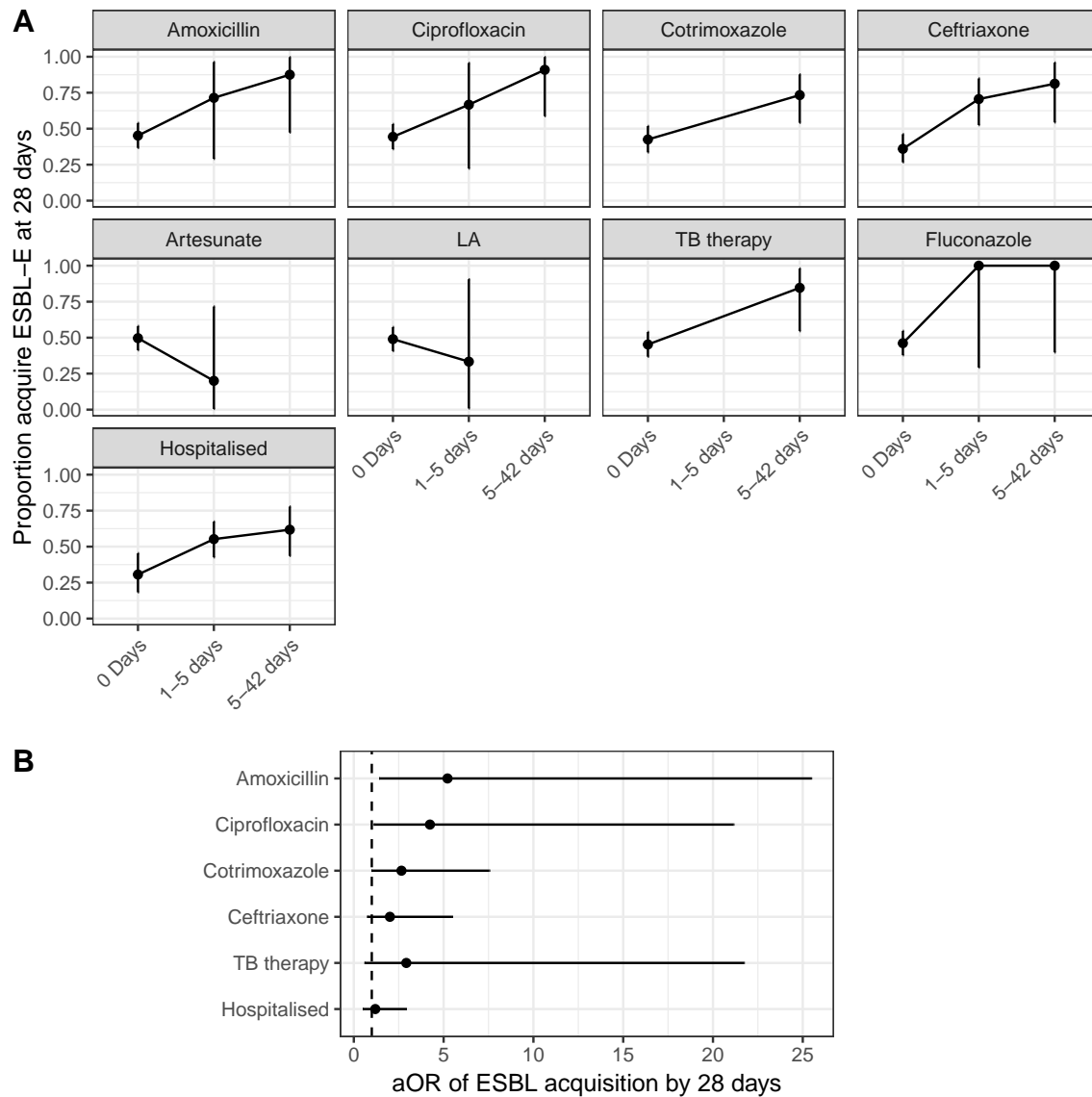


Figure 5.7: Univariable (A) and multivariable (B) associations of antimicrobial and hospital exposure with acquisition of ESBL-E by 28 days. A: These plots show the proportion of participants who have no detectable ESBL-E baseline but who do at 28 days, as a function of various exposures. All antibacterials and hospitalisation show an association between exposure and ESBL-E acquisition, with a suggestion of a dose-response relationship, though confidence intervals are wide in many cases. Antimalarials show no apparent relationship though, as with fluconazole, the wide confidence intervals make it difficult to draw any conclusions. The results from logistic regression to predict ESBL-E acquisition are shown in (B); collinearity and small dataset size means that confidence intervals are so large as to make the model useless.

## 5.5 Discussion

In this chapter, I have presented the data which will be used to address the second aim of this thesis - to describe, and identify determinants of, ESBL-E acquisition and carriage in Malawian adults. It is possible to draw several conclusions from these data. First, community ESBL-E carriage in Blantyre is common. The baseline community carriage prevalence of 28% is considerably higher than the 4-7% seen in Europe[165–168] and comparable to the sSA pooled community prevalence estimate presented in Chapter 1 of 18% (95% CI [11-27%]).

The associations of baseline ESBL-E carriage give insight into the routes of community ESBL-E transmission in the urban Malawian setting. The high community prevalence without apparent healthcare contact suggests that community transmission is common. Household crowding and use of unprotected water sources are associated with ESBL-E colonisation, suggesting both household person to person and environmental transmission routes are of relevance. The number of adults in the household, rather than the number of children, was associated with ESBL-E carriage in this study - which recruited only adults - suggesting that within the household adult to adult transmission is a more important route than child to adult transmission. This could be for a number of reasons - if the ESBL-E prevalence were low in children, for example. Though children were not sampled in this study and so the data here can not address that hypothesis, data from other studies suggest this is unlikely: community prevalence in children ranged from 10-59% in four studies in the Central African Republic[250], Senegal[270] and Tanzania[248,255], and is hence comparable to the adult community prevalence seen in this study. Behavioural factors, or a lower bacillary burden in children could also account for the associations seen here. The seasonality of ESBL-E carriage prevalence is also consistent with an environmental transmission route of ESBL-E - environmental faecal and hence ESBL-E contamination would be likely higher in the rainy season - but behavioural factors (e.g. more indoor crowding during the rains) or other causal pathways (e.g. more febrile illness and hence hospitalisation and/or antimicrobial exposure) should also be considered as possibilities.

These data also offer insight into the role of HIV as a driver of ESBL-E colonisation. Prior to this study, only two studies in sSA had assessed associations of HIV and community ESBL-E carriage: one, in pregnant women, found no association[259], and one, in healthy children under two, found a strong association[255]. HIV is known to have a profound effect on gut function[328] and it is conceivable that a direct HIV effect on e.g. the microbiota could result in reduced colonisation resistance to ESBL-E, or that increased antimicrobial or healthcare exposure in the HIV infected could result in a higher colonisation prevalence. However, in the Malawian adults considered here, there is a clear association between HIV-infection and ESBL-E carriage which seems to be largely mediated by CPT. Given that the adult HIV prevalence in Malawi is estimated to be 9.6%[329] and Malawian HIV

guidelines mandate lifelong CPT for people living with HIV[330], CPT is potentially a significant driver of ESBL-E colonisation in this setting.

It is also clear that there are significant associations between antimicrobial exposure, hospitalisation, and ESBL-E carriage prevalence, though the analysis presented here is unable to provide an insight into the relevant importance of hospitalisation versus antimicrobial exposure. Prior hospitalisation was strongly associated with baseline ESBL-E carriage on multivariable logistic regression, but prior antimicrobial exposure (apart from CPT) was not, though with confidence intervals that include a clinically relevant effect size. A better estimate of the relative effects can come from the longitudinal sampling data. The study was designed to explore the relative effects of hospital exposure and antimicrobial exposure on ESBL-E colonisation, and so to produce groups with different levels of these exposures. Despite some crossover of exposure, this was largely a success, with much reduced antimicrobial exposure in Arm 2 compared to Arm 1, and virtually no antimicrobial or hospital exposure in Arm 3. In an unadjusted analysis, Arm 1 participants show a greater increase in ESBL-E carriage prevalence than Arm 2 participants, suggesting that antimicrobial exposure, rather than hospitalisation, is driving apparent acquisition. The prevalence of ESBL-E colonisation following antimicrobial exposure is striking, with 78% of Arm 1 participants colonised by the day 7 visit. A return to pre-admission prevalence is apparent over the 6-month study period.

However this conclusion - that antimicrobial exposure rather than hospitalisation is driving apparent ESBL-E acquisition - is very open to bias from confounding, because antimicrobial-unexposed participants are different from antimicrobial unexposed: younger, less likely to be HIV-infected, less CPT exposure, and a shorter length of hospital stay. A multivariable approach is necessary to provide unbiased estimates of the effects of antimicrobial use and hospitalisation, but collinearity of exposures and a small dataset size means that a simple modelling approach - logistic regression - produces such uncertain effect sizes as to be useless. A different modelling approach, using continuous time Markov models, is presented in Chapter 8.

The overall trend of ESBL-E carriage at any time point in the study obscures the complex within-individual picture, with a pattern of multiple transitions between the detectable and undetectable ESBL-E states over the study period for many individuals. There could be several reasons for this: if a participant is truly continually colonised, then intermittent shedding of ESBL-E or inadequate test sensitivity could explain apparent lack of ESBL-E in culture. There are clearly some participants who are continually colonised except for at one or two time points, which is suggestive of one or both of these scenarios. Alternatively, these patterns could represent true acquisition or loss events.

The ESBL-E cultured from the participants in this study are, as expected, largely *E. coli* and, to a lesser extent *K. pneumoniae*, as seen in other studies in sSA (see Chapter 1),

but also in high income settings such as Europe[177]. The prevalence of amikacin and meropenem resistance was very low, although was present. Meropenem was introduced to the Malawian national formulary in 2015[331], and is only sporadically available in QECH. Previous studies of antimicrobial resistance in Blantyre have not found any genotypic or phenotypic carbapenem resistance, though the MLW laboratories do not routinely test for carbapenem sensitivity. To my knowledge the carbapenem resistant isolates found in this study are the first to be described in Malawi, and the mechanisms of carbapenem resistance are further explored in Chapter 6. Cotrimoxazole resistance was near universal, which may be related to widespread CPT use. Interestingly, many of the isolated ESBL-E were sensitive to chloramphenicol. This drug was, prior to the introduction of ceftriaxone, first line treatment (along with penicillin) for sepsis in Malawi, but was replaced due to a poor side effect profile and the ease of administration of once-daily ceftriaxone. Clearly these isolates are carriage rather than invasive isolates and so may have different sensitivity patterns to the bacteria that cause invasive infections, but this raises the prospect that chloramphenicol might be reintroduced as a reserve antibiotic in the treatment of ESBL-E in Malawi.

### 5.5.1 Limitations

There are limitations to this study and analysis. There was significant drop-out of Arm 1 and 2 participant through the study period. The proportion of participants truly lost to follow up or voluntarily withdrawn from the study was low, however, and most withdrawals were due to death or transfer out, which reflect the study populations of interest in urban Blantyre. Logistic difficulties - participants being unable to travel to deliver a sample, or being unavailable for a home visit - resulted in failure to collect 13% of eligible sample, and again reflects the challenges inherent in the setting. Though the study protocol mandated the timing of visits, the actual visit times have a broad temporal distribution. Nevertheless, there does not seem to be any systematic bias in the missingness of participants or samples, and if that is the case then the conclusions drawn should stand. HIV testing was not carried out on community members, and so there is a high proportion of arm 3 participants with an unknown HIV status, and the status that is recorded is self reported, which could result in misclassification of HIV status and hence bias. Indeed, all baseline exposures were self-reported and not verified, which could result in bias.

## 5.6 Conclusions and further work

ESBL-E colonisation is common in adults in Blantyre, and ongoing community transmission with both person to person and environmental transmission routes seems likely. Rapid

apparent acquisition of ESBL-E occurs in hospitalised participants exposed to antimicrobials with a return to pre-admission ESBL-E colonisation prevalence by six months. The aggregate data conceals a complex picture within individuals, however, with multiple apparent transitions between the colonised and uncolonised state for many individuals over the study period. The relative contribution of hospitalisation and antimicrobial exposure in driving apparent acquisition events is not clear and logistic regression models failed to provide any insight.

Several questions arise from this analysis and form the basis of further work in this thesis.

1. What is the mechanism underlying the frequent state transitions (from colonised to uncolonised and back again) for many participants in this study?
2. Are these true acquisitions and losses, intermittent shedding, or a failure to detect ESBL-E which are present?
3. What are the relative contributions of hospitalisation versus antimicrobial exposure in driving the sharp increase in ESBL-E carriage prevalence following hospital admission and antimicrobial exposure? Specifically, what is the biological mechanism of this apparent increase? Is it associated primarily with hospitalisation - which could represent a true ESBL-E acquisition from a contaminated hospital environment - or primarily with antimicrobial exposure - which could be an enrichment of already-carried ESBL-E which was undetected at baseline - or a synergistic combination of the two?

I will spend the next three chapters attempting to address these questions, using whole genome sequencing as a high resolution bacterial typing tool to track bacteria within participants, and continuous-time Markov models to model state transitions between colonised and uncolonised states over time.

## Chapter 6

# The genomic landscape of ESBL producing *E. coli* in Blantyre, Malawi

### 6.1 Chapter overview

In this chapter I present the results of whole-genome sequencing of 473 *E. coli* isolates, cultured from 230 participants at a median of 2 (IQR 1-5) time points per participant. These represent one colony pick from all stool samples which cultured *E. coli* up to the time of shipping in October 2018. The isolates were largely phylogroup A (43% [204/473]), and ST131 was the most common ST (14% [64/473]). A global collection of 800 *E. coli* genomes was used to put the isolates from this study in a global context and a maximum likelihood phylogenetic tree constructed. Isolates from this study were largely spread throughout the tree, but there were exceptions: Malawian ST410 and ST167 were monophyletic suggesting perhaps recent introduction to Malawi. These are recognised emerging high-risk clones. These findings suggest that there no obvious restrictions in mixing of *E. coli* between Blantyre and the rest of the world.

A diverse range on AMR genes were identified, with a median 16 (IQR 12-17) per isolate. ESBL genes were largely of the *bla*<sub>CTXM</sub> family, and dominated by *bla*<sub>CTXM-15</sub>. The abundance of genes seems to reflect local antibiotic pressures: genes conferring cotrimoxazole and aminoglycoside resistance were almost ubiquitous, and quinolone, chloramphenicol and tetracycline resistance genes also frequently seen. Carbapenem antibiotics are rarely available, and carbapenem resistance genes rare. One carbapenemase, a *bla*<sub>NDM-5</sub>, was identified, on an IncX plasmid very similar to one first identified in India around 2011. To my knowledge the first carbapenemase described in Malawi. Its identification despite

little antibiotic pressure along with the apparent ease international of spread of *E. coli* suggested by this analysis suggest widespread resistance could rapidly emerge following wider carbapenem roll out. Antibiotic stewardship interventions that can balance access to and restriction of last line antibiotics in low resource settings are urgently needed.

## 6.2 Introduction and chapter aims

The previous chapter presented an overview of the longitudinal ESBL-E sampling that was undertaken in this study. It is the overall aim of this thesis to use whole genome sequencing (WGS) as a high-resolution typing method to longitudinally track bacteria within participants. In this chapter, however, I present a descriptive analysis of the genomes of 473 *E. coli* isolates sequenced for this study, before moving on to this specific aim in the following chapter. The aims of the analysis presented in this chapter are, therefore:

1. Perform quality assurance and control of the sequenced genomes for downstream analysis.
2. Place the isolates from this study in a global context using phylogenetics and a global *E. coli* collection comprised of available genomes.
3. Provide a description of the AMR determinants in the sequenced *E. coli* isolates.

## 6.3 Methods

### 6.3.1 Bioinformatic pipeline

One *E. coli* colony from each patient sample was picked for DNA extraction and 150bp paired-end short-read whole genome sequencing (WGS) using Illumina HiSeq X10 at the Wellcome Sanger Institute (WSI). Read quality control was undertaken with Kraken v0.10.6 and Braken v1.0 to assign reads to species[332] and the WSI QC pipeline which maps a random 100 Mbases from each sample to a reference and calculates depth of coverage, number of heterogeneous SNPs, GC content and insert size. Samples that contained > 80% non *E. coli* reads were discarded the remainder were *de novo* assembled with SPAdes v3.11.0[333]. Assembly statistics (e.g. assembly length, N50) were calculated with QUAST v4.6.0[334] and completeness and contamination of the assemblies assessed by checkM v1.0.7[335], which uses presence or absence of lineage-associated collocated ubiquitous single-copy genes to estimate these quantities. Contaminated assemblies (with checkM-defined contamination of > 25%) or poor assemblies (with less than 1Mb assembled length) were discarded. Annotation was carried out with prokka v1.5[336] with a genus specific database from RefSeq and the Roary v1.007 pan-genome pipeline[337] was used to identify



a core genome, considering genes contained in > 99% of isolates to be core. A core gene multiple sequence alignment was generated using maaft v7.205[338], SNP-sites identified using SNP-sites v2.4.1[339] and the resultant SNP alignment used to build a maximum likelihood phylogenetic tree using IQ-TREE v1.6.3[340], using ascertainment bias correction to correct for the fact that the input pseudosequence contained only variable sites, and using the ModelFinder module used to find the best fitting nucleotide substitution model. This calculates the likelihood of a number of different models and chooses the model with the highest (best fitting) Bayesian Information Criterion, a statistic which penalises model parameters. Reliability of inferred branch partitions was assessed with 1000 bootstrap replicates. Trees were visualised in the ggtree v1.14.4 package[341] in R.

ARIBA v2.12.1[342] was used to identify AMR-associated genes using the SRST2 database[343], to identify plasmid replicons using the PlasmidFinder database[344] and to perform *in silico* multi-locus sequence typing (MLST) using the database from <http://mlst.warwick.ac.uk/mlst/dbs/Ecoli> accessed via [www.pubmlst.org](http://www.pubmlst.org). The  $\beta$ -lactamase genes *ampC1*, *ampC2* and *ampH* were excluded from the analysis of AMR determinants as they do not usually cause a resistant phenotype in *E. coli*. Because quinolone resistance often results from SNPs in the chromosome in the quinolone resistance determining regions (QRDRs) of the *gyrA*, *gyrB*, *parE* and *parC* genes - rather than acquisition of whole AMR-determining genes, as is the case with the other genes sought by ARIBA - these genes were downloaded from the comprehensive antimicrobial resistance database (CARD, <https://card.mcmaster.ca/>) and ARIBA used to call SNPs in them, with default settings. *E. coli* phylogrouping was performed with a quadruplex *in silico* PCR using the Clermont scheme[345] and *isPcr* v33x2 (<https://github.com/bowhan/kent/tree/master/src/isPcr>)

### 6.3.2 Global *E. coli* collection

In order to place the isolates from this study in a global context, published *E. coli* assemblies were downloaded from the WSI servers. These included 149 ESBL-producing *E. coli* from a single centre study in Chachoengsao province, eastern Thailand[346]. In this study, human clinical isolates from standard care in Bhuddhasothorn hospital were selected on the basis of the ESBL phenotype, and environmental samples were collected as part of a cross sectional study and selectively cultured for ESBL-E in 2014-2015. I also downloaded assemblies of 362 enterotoxigenic *E. coli* (ETEC), selected for an ETEC genomic study from the Gothenburg University ETEC collection to represent a broad collection of ETEC isolated worldwide from 1980-2011[347]; 185 atypical enteropathogenic *E. coli* (aEPEC) sequenced for a study of aEPEC and selected from samples from the Global Enteric Multicentre Study (GEMS) in seven centres in Africa and Asia between 2007-2011[348]; and 94 *E. coli* from QECH in Blantyre, Malawi, a combination of invasive (bloodstream and CSF) and carriage isolates, selected for diversity in AMR phenotype from 1996-2014[279]. Details of

the included samples are given in the appendix to this chapter.

Phylogroup and MLST were determined for these context genomes as described above. AMR genes were identified with Ariba and the SRST2 database, as above, and context genomes were classified as ESBL if they contained any Group 2be ESBL gene (see Bush-Jacoby scheme in Chapter 1).

### 6.3.3 Statistical analysis

Association of AMR genes with phenotype was expressed as odds ratios and tests of association used Fisher’s exact test. In order to explore clustering of AMR genes, the Jaccard index was calculated for a given AMR-gene pair using the *philentropy* v0.3.0 package in R. The Jaccard index, a measure of the similarity of two sets of data, is defined as *intersection over union*; in this context, for a given pair of AMR genes  $x$  and  $y$ , the Jaccard index  $J(x, y)$  is the number of isolates that contain both gene  $x$  and  $y$  divided by the total number that contain either  $x$  or  $y$ . By definition it lies between 0 ( $x$  and  $y$  never co-occur) and 1 ( $x$  and  $y$  always co-occur). Co-occurrence matrices using the Jaccard index were plotted using the *pheatmap* v1.0.12 package in R. The statistical significance of co-occurrence of genes was assessed by generating 2x2 contingency tables for a given gene pair and p values generated using a Fisher’s test with Bonferroni correction; a p value of less than 0.05 was considered statistically significant. Co-occurrence networks of genes occurring commonly together (defined as Jaccard index  $> 0.5$ ) at a rate greater than expected by chance ( $p < 0.05$  following Bonferroni correction) were plotted using *igraph* v1.2.2[349] and *ggraph* v1.0.2 in R.

## 6.4 Results

### 6.4.1 Samples and quality assurance and control

There is a detailed description of microbiological procedures in Chapter 2. In total, 519 *E. coli* underwent DNA extraction and were shipped from Malawi to WSI; these represented all sequential isolates at the time of final DNA extraction, which occurred in two batches in February 2018 and October 2018. Kracken/Bracken read assignment of these samples is shown in Figure 6.1. The majority of samples have  $> 90\%$  of reads assigned to *E. coli*; a minority have  $< 90\%$  of reads assigned to *E. coli* but a very closely related species such as *Shigella*, and as such are likely to be pure *E. coli* culture with read misclassification. However, 12 samples have  $> 80\%$  reads assigned to a non- *E. coli* species such as *Klebsiella pneumoniae*. These samples were assumed to represent upstream species misidentification, and were excluded. There also exists the possibility of within-participant transcription

error. In the freezer archive, all samples from a single participant at a single time point have the same sample ID, making an error possible; by definition (as only one *E. coli* was sequenced at any time point in any individual) this would result in a species substitution. Samples from different time points and different participants are clearly demarcated however, making between-time point or between-participant errors very unlikely. Any such error would therefore result in a sample being identified as non- *E. coli* and exclusion from the analysis, reducing power but not introducing bias.

Of the remaining 507 samples, there were a median (IQR) of  $2.3 \times 10^6$  ( $2.1 - 2.5 \times 10^6$ ) reads, with a median (IQR) depth of coverage (obtained by mapping a random 100Mbases to a reference *E. coli* genome, *Escherichia coli* strain K-12 substrain MG1655, NCBI reference NC\_000913.3) of 58 (51-66). One sample had an order of magnitude lower number of reads ( $2.9 \times 10^5$ ) with depth of coverage 0; this was assumed to represent sequencing failure and it was excluded from further analysis.

The output from quast and checkM are shown in Figure 6.2, where N50 (the minimum contig length upon which at least half assembled bases are contained) is plotted as a function of total assembled length. The expected *E. coli* genome length is around 4.6Mb and most samples cluster close to this at a total assembled length of  $\sim 5$ Mb. However it is clear that some assemblies have failed, with low N50 and low assembled length. It is also apparent that some samples seem to be contaminated, as indicated by low N50 and much longer than expected total assembled length. Defining assembly failure as  $< 1$ Mb assembled length (triangles in the plot,  $n = 9$ ) and contamination as checkM-defined contamination of  $> 25\%$  (blue points in the plot,  $n = 24$ ) and excluding both groups results in 33 further samples being excluded from further analysis.

In total, therefore, 46/519 (9%) of samples which were submitted for sequencing were excluded from downstream analysis. The remaining 473 samples represent 69% (474/686) of the cultured *E. coli* in this study, and were recovered from 230 participants. 354 are from patients with sepsis, 86 are from hospitalised inpatients and 33 are from community members, with a median of 2 (range 1-5) samples per participant.

#### 6.4.2 Phylogroup, MLST and core genome phylogeny of study isolates

The commonest *E. coli* phylogroup was phylogroup A, followed by phylogroup B2, F, B1 and C and D (Table 6.1). Two samples were Clade I or II (so called cryptic clades) and six were unknown phylogroup using the Clermont PCR scheme. In the MLST analysis, 56 recognised sequence types (STs) were identified, and 12 samples were novel STs; however over half (249/473 [53%]) of samples were represented by the top seven most frequent STs (Figure 6.3). ST131 was the most commonly isolated ST (64/473 [14%] of isolates) followed by ST410 (45/473 [10%] of isolates) and ST167 (38/473 [8%] of isolates).

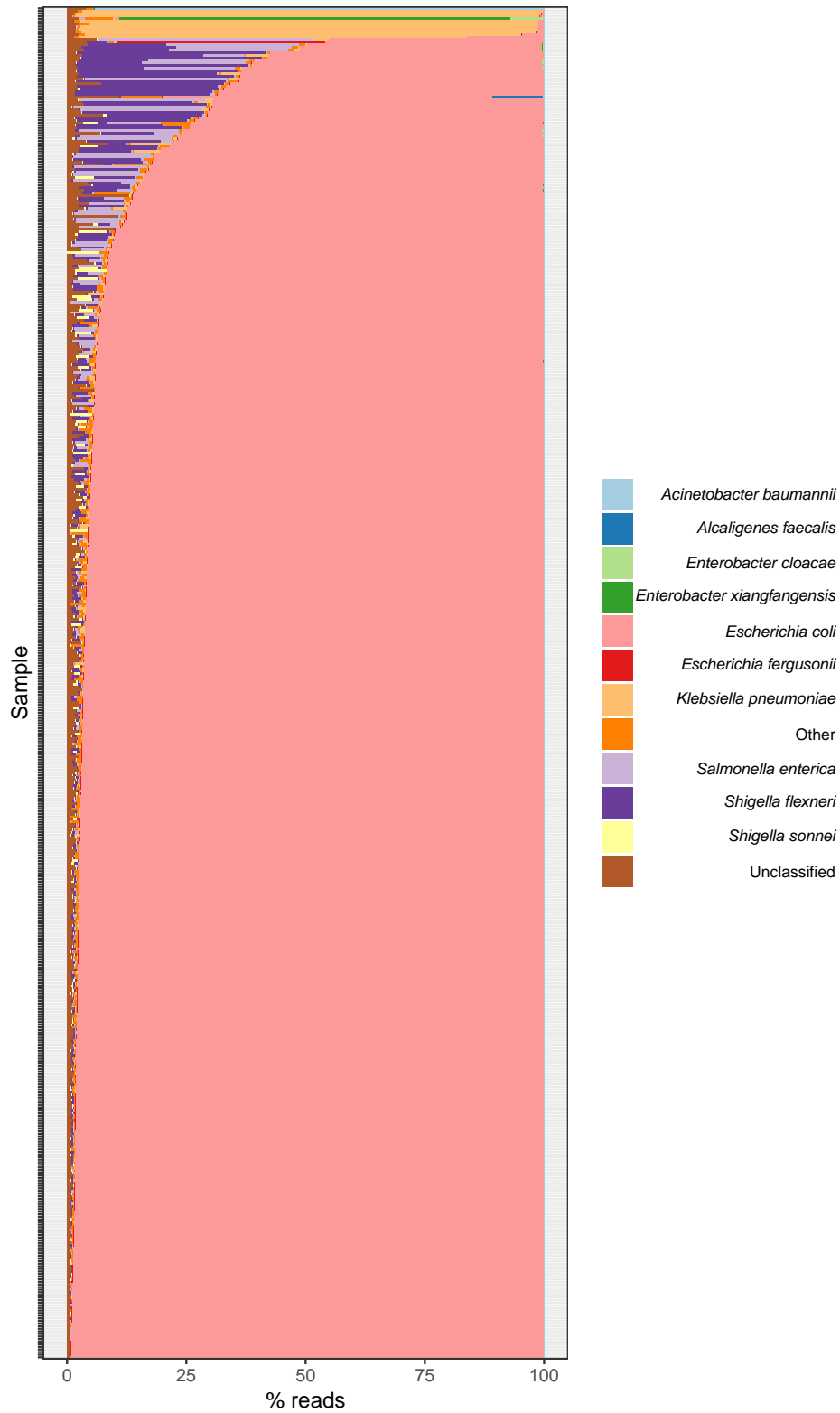


Figure 6.1: Species read assignment of all samples. Each horizontal bar is one sample. Most samples have > 90% reads assigned to *E. coli* or related species (e.g. *Shigella* spp.) but twelve samples are likely not *E. coli* and excluded from further analysis.

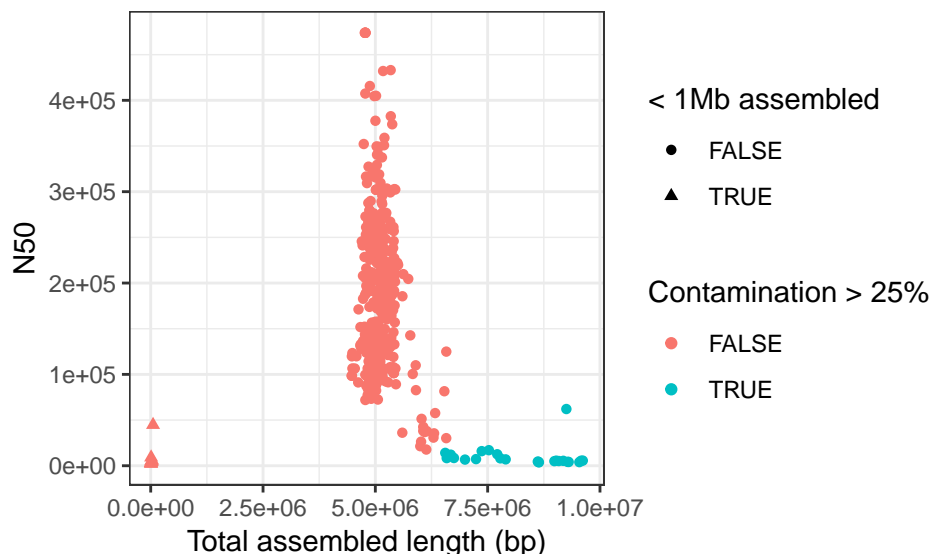


Figure 6.2: N50 as a function of total assembled length. Failed assemblies with less than 1Mb assembled shown as triangles. Contaminated assemblies with checkM-defined contamination above 25% shown in blue. Both of these groups of assemblies were excluded from further analysis.

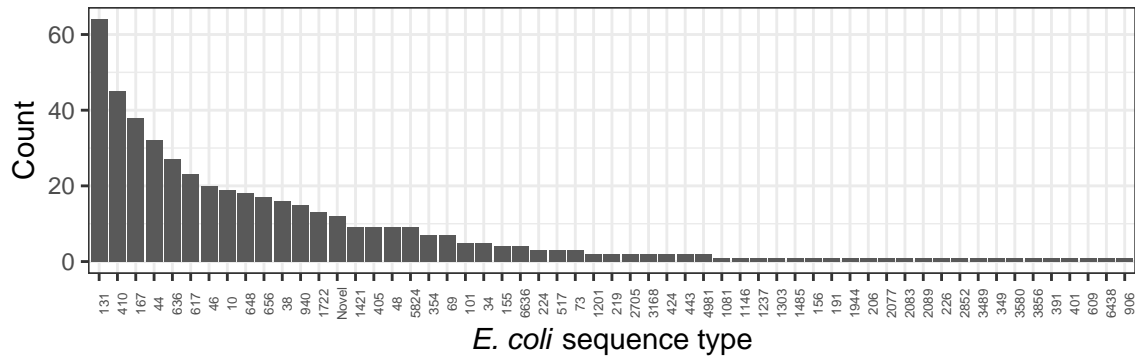
Table 6.1: Phylogroup distribution of sequenced *E. coli* isolates.

Phylogroup	n
A	204/473 (43%)
B2	96/473 (20%)
F	53/473 (11%)
B1	43/473 (9%)
C	43/473 (9%)
D	26/473 (5%)
Unknown	6/473 (1%)
Clade I or II	2/473 (0%)

The Roary pan-genome pipeline identified a core genome in the study isolates of 2966 genes, with a pan-genome of 26,840 genes. The resultant core gene pseudosequence of length 1,388,742 bases contained 99,693 variable sites, which were used to infer the maximum likelihood phylogenetic tree. The IQTREE ModelFinder module determined that a general time reversible (GTR) model with FreeRate site heterogeneity with 5 parameters provided the best fit to the data. The inferred tree is shown in Figure 6.4 along with isolate phylogroup and ST; in general, as expected, STs were largely monophyletic and phylogroups tended to cluster together.

### 6.4.3 Study isolates in a global context

The global collection of *E. coli* comprised 1273 samples, including the 473 from this study. 753/1253 (60%) were from Africa, 335/1253 (27%) from Asia and 167 (13%) from South America. The majority of samples, 1026/1253 (82%), were from stool, with 106/1253 (8%) truly invasive samples from blood or CSF and 63/1253 (5%) possibly invasive samples

Figure 6.3: *E. coli* multilocus sequence type distribution

from urine, pus, or sputum. 65/1253 (5%) samples were environmental, all from Thailand. 670/1253 (53%) of samples contained at least one ESBL-encoding gene. The majority of isolates with an ESBL gene (622/670 [92%]) came from this study or the Thai ESBL study, a potential source of bias. Phylogroup A was the commonest phylogroup in the global collection (482/1273 [38%]), followed by B1 (333/1273 [26%]) and B2 (191/1273 [15%]); phylogroup C was uncommon in the global collection (74/1273 [6%]) but the majority of the phylogroup C samples came from this study (43/74 [58%]). All of these 43 phylogroup C isolates belonged to a single ST, ST410; this ST was not seen at all in the previous Malawian study of largely invasive isolates, despite being the second-commonest ST in this study, and was unusual in the global collection (11/800 [1%] ST410 in global collection vs 43/473 [9%] in this study). Similarly, the third-commonest ST in this study, ST167, was not seen at all in the global collection. However, ST131, the commonest ST in this study, was again the commonest ST in the global collection.

The Roary pan-genome pipeline identified 2872 core genes in a pan genome of 44,840 genes; this large pan-genome is consistent with the open *E. coli* pan genome that will continue to increase in size as isolates are added. The core gene alignment contained 604,817 bases with 77,194 variable sites, which were used to infer the maximum likelihood phylogenetic tree, using the same nucleotide substitution model as previously.

The phylogeny is reconstructed in Figure 6.5. Isolates from this study are distributed throughout the tree, and there is widespread mixing of isolates from diverse geographic regions. Though invasive isolates are spread throughout the tree, there is a tendency for them to cluster together, particularly in phylogroup B2, a phylogroup with has a recognised association with ExPEC[350]. The Malawian ST410 and ST167 isolates clustered tightly together, but by comparison, ST131 isolates from this study were distributed among ST131 isolates from other studies, both in Malawi and elsewhere (Figure 6.6).

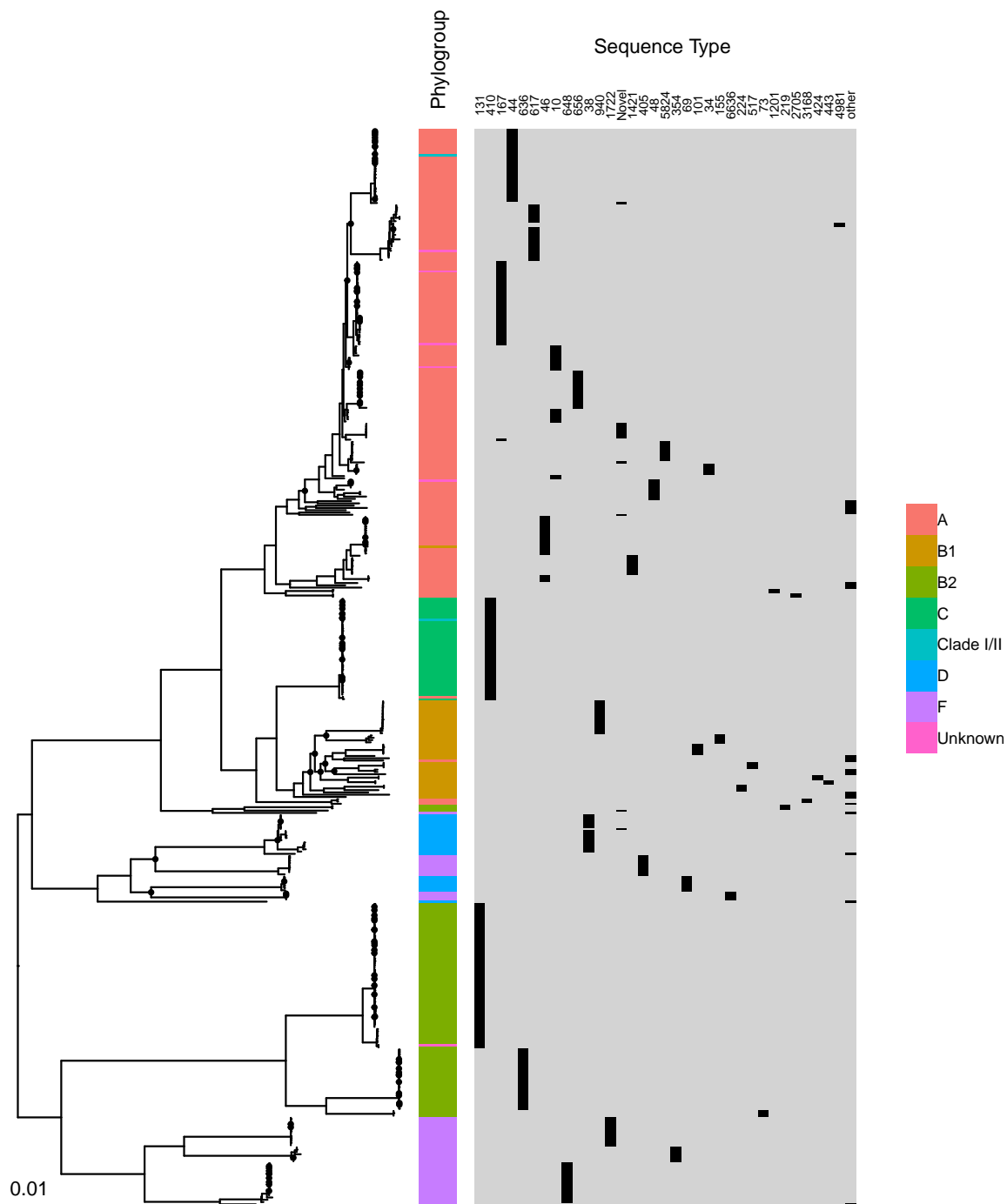


Figure 6.4: Midpoint rooted maximum likelihood phylogenetic tree of included study *E. coli* isolates showing phylogroups and sequence types. Bootstrap support of less than 90% is indicated by a black circle at a given node. Scale bar indicates 0.01 SNPs/site.

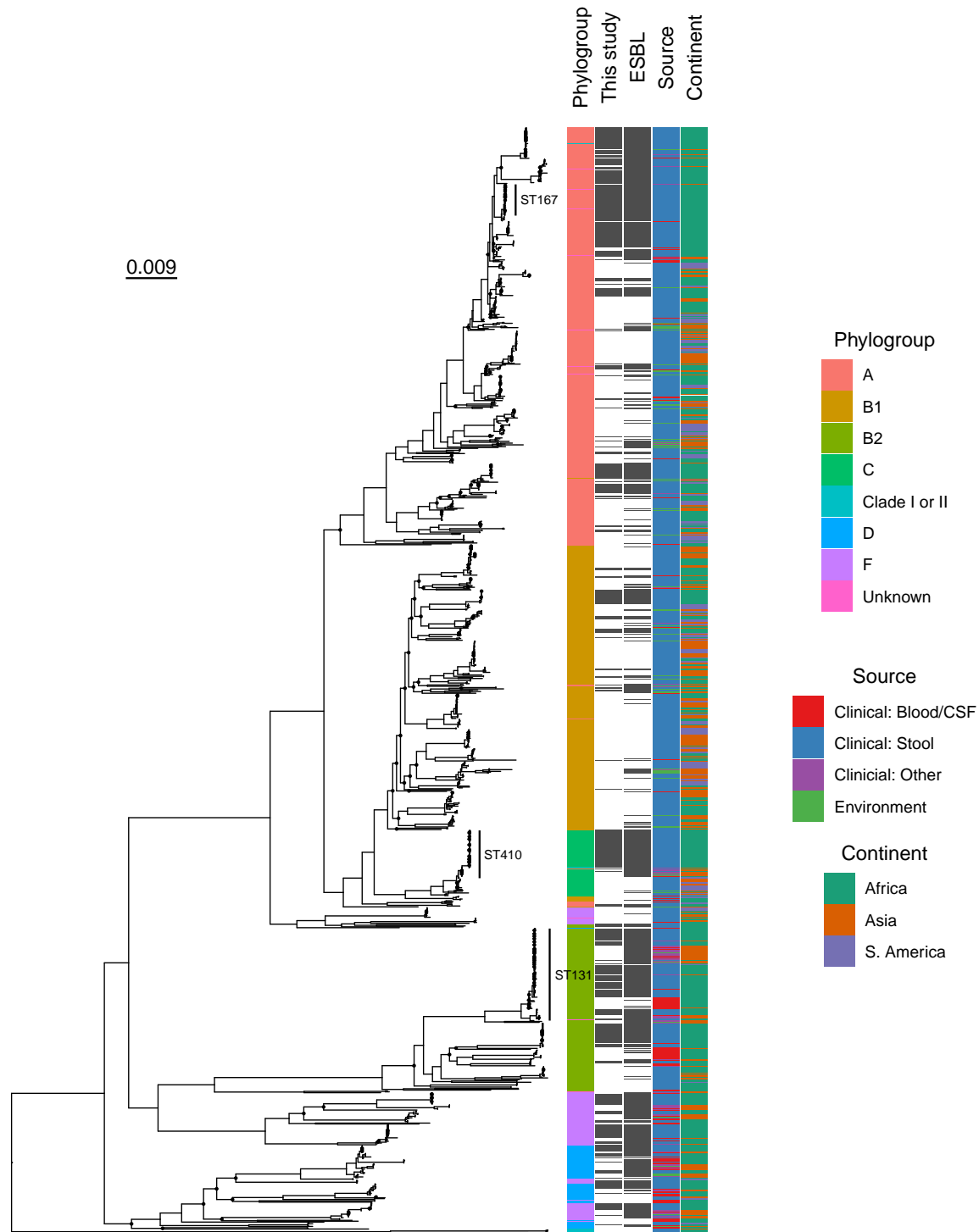


Figure 6.5: Midpoint rooted maximum likelihood phylogenetic tree of included study *E. coli* placed in the context of a global collection of genomes, showing phylogroups, source sample type and continent of isolation (coloured bars). Dark grey bars indicate isolates from this study or isolates with ESBL gene presence, as labelled (this study or ESBL, respectively). Three most frequently isolated STs in the current study (131, 410 and 167) labelled. Bootstrap support of less than 90% is indicated by a black circle at a given node. Scale bar indicates 0.009 SNPs/site.





Figure 6.6: Subtree of ST410 (A, left) and ST131 (B, right), and ST167 (C, bottom) showing multiple introductions of ST131 into Malawi, in comparison to a single introduction of clonal ST410 and ST167 clades. Colour of tree tip label indicates isolation from this study (red) or other studies (black), and coloured heatmap indicates country of isolation. Note that the scale bar in C is an order of magnitude different from A and B.

#### 6.4.4 Antimicrobial resistance determinants

All identified AMR genes are shown in Figure 6.7A, alongside a summary of number of isolates with resistance mutations to given antimicrobial classes (Figure 6.7B) and the phenotypic resistance of the isolates for which phenotypic antimicrobial resistance testing was carried out (449/473 [95%]). The isolates contained a median (IQR) of 16 (12-17) resistance genes, and 100 different resistance alleles were identified in total. A description of resistance gene by class is given below. In this section I consider and present data only from the isolates from this study, rather than data from the context genomes presented above.

##### 6.4.4.1 $\beta$ -lactam resistance

All isolates contained at least one gene that conferred resistance to third-generation cephalosporins, either an ESBL gene ( $n=472$ ) or a carbapenemase ( $n=1$ ). The majority of ESBL-gene containing isolates contained only one ESBL gene (459/472 [97%]); fewer contained 2 (13/472 [3%]) and none contained more than 2. The *bla*<sub>CTX-M</sub> family was most commonly represented in this collection, and 319/473 (67%) of isolates contained *bla*<sub>CTXM-15</sub>. (26/473 [5%] of isolates) contained *bla*<sub>SHV</sub> genes. The  $\beta$ -lactamases *bla*<sub>TEM</sub> and *bla*<sub>OXA</sub> were very common, most commonly the penicillinases *bla*<sub>OXA-1</sub> and *bla*<sub>TEM-1</sub> in 186/473 [39%] and 289/473 [61%] of isolates respectively. Plasmid-mediated *bla*<sub>ampC</sub> genes were identified in 45/473 (9%) of isolates, almost all (44/45) *bla*<sub>CMY-94</sub>, which was lineage-restricted to the ST410 isolates. Presence of *bla*<sub>ampC</sub> is unexpected as all of these isolates were confirmed to be ESBL-producers by combination disc testing. This testing uses cephalosporin-containing discs both with and without clavulanic acid, and confirms EBSL production by a difference in zone size between these discs, as ESBL enzymes are inactivated by clavulanic acid. However, the cephalosporins used in this test are likely to be hydrolysed by *ampC* enzymes, and if these isolates were producing such enzymes it could confer cephalosporin resistance regardless of the presence or absence of clavulanic acid. This was not the case for any of these isolates; none of them hydrolysed the cephalosporins used in the presence of clavulanic acid. It may be that the *bla*<sub>CMY</sub> genes were not expressed or have been inactivated in the ST410 clade.

The single carbapenemase gene identified was s *bla*<sub>NDM-5</sub>; the isolate harbouring this gene was recovered from the stool of a 67-year old man with no history of foreign travel nor hospitalisation. He had been admitted to the hospital with fever seven days previously and treated with seven days of intravenous ceftriaxone for sepsis, the source of which was not clear. He made an uneventful recovery, and no carbapenemase-containing isolate was recovered from his stool at any other time. The *bla*<sub>NDM-5</sub> gene was carried on a partially assembled IncX3 plasmid. BLAST of this assembly against the NCBI database showed

that this contig had 99% sequence identity with a previously sequenced pNDM-MGR194 46.2 kbp blaNDM-5 containing Inc-X3 plasmid found in India between 2011-13[351]. We fully assembled the plasmid by mapping reads back to pNDM-MGR194 with Burrows-Wheeler alignment and found it to be extremely similar, with only 13 SNPs compared to pNDM-MGR194.

#### 6.4.4.2 Quinolone resistance

Quinolone resistance can be mediated by a number of different mechanisms and at least one known determinant of quinolone resistance was present in 88% (418/473) isolates. 108/473 (23%) of isolates contained one of the the plasmid-mediated quinolone resistance (PMQR) genes *qnrS*, *qnrB* or *qepA*; 172/473 (36%) carried the acetyltransferase *aac(6')-Ib-cr* which can hydrolyse fluoroquinolones with an amino nitrogen on the piperazinyl ring, such as ciprofloxacin and norfloxacin[352]. Nonsynonymous mutations were identified in at least one of the DNA gyrase subunits *gyrA* or *gyrB* or topoisomerase IV subunits *parC* or *parE* in 349/449 (78%) of isolates. The majority of these mutations were well-described QRDR mutations (codon 83 and 87 in *gyrA*, codon 80 and 84 in *parC* and codon 458 in *parE*[353], Figure 6.8A), and they tended to cluster together (Figure 6.8B). Mutations in *gyrB* were very unusual, with only 3 identified in the dataset.

The association between phenotypic and genotypic resistance is complex but some patterns seem clear (Figure 6.8B). The well recognised constellation of mutations in *gyrA* at codon 83 and 87 and codon 80 of *parC* were strongly associated with phenotypic ciprofloxacin resistance (OR 7.3 [95% CI 4.4-12.4],  $p < 0.001$  by Fisher's exact test), but *gyrA* codon 83 mutations alone seemed insufficient: of the 63 samples with this mutation alone and available AST data only 30% (19/63) showed phenotypic resistance. Similarly, presence of the *qnrS* plasmid-mediated gene seemed insufficient alone to confer phenotypic resistance: of 57 samples in which this gene was present and AST data were available, only 25/57 (43%) had phenotypic ciprofloxacin resistance by disc diffusion testing.

#### 6.4.4.3 Aminoglycoside resistance

The identified aminoglycoside resistance determinants are shown in Figure 6.9. Most aminoglycoside resistance genes are classified into three families based on their ability to acetylate, phosphorylate, or adenylate amino or hydroxyl groups found at various positions around the aminoglycoside molecule, and are called acetyltransferases (AACs), nucleotidyltransferases (ANTs), or phosphotransferases (APHs)[354]. They usually further identified by the site of action in terms of the aminoglycoside carbon atom upon which they act, subclass, and individual identifier. For example, the enzyme AAC(3)-Ib (gene *aac(3)-Ib*) refers to an acetyltransferases acting at position 3 of subclass I and individual

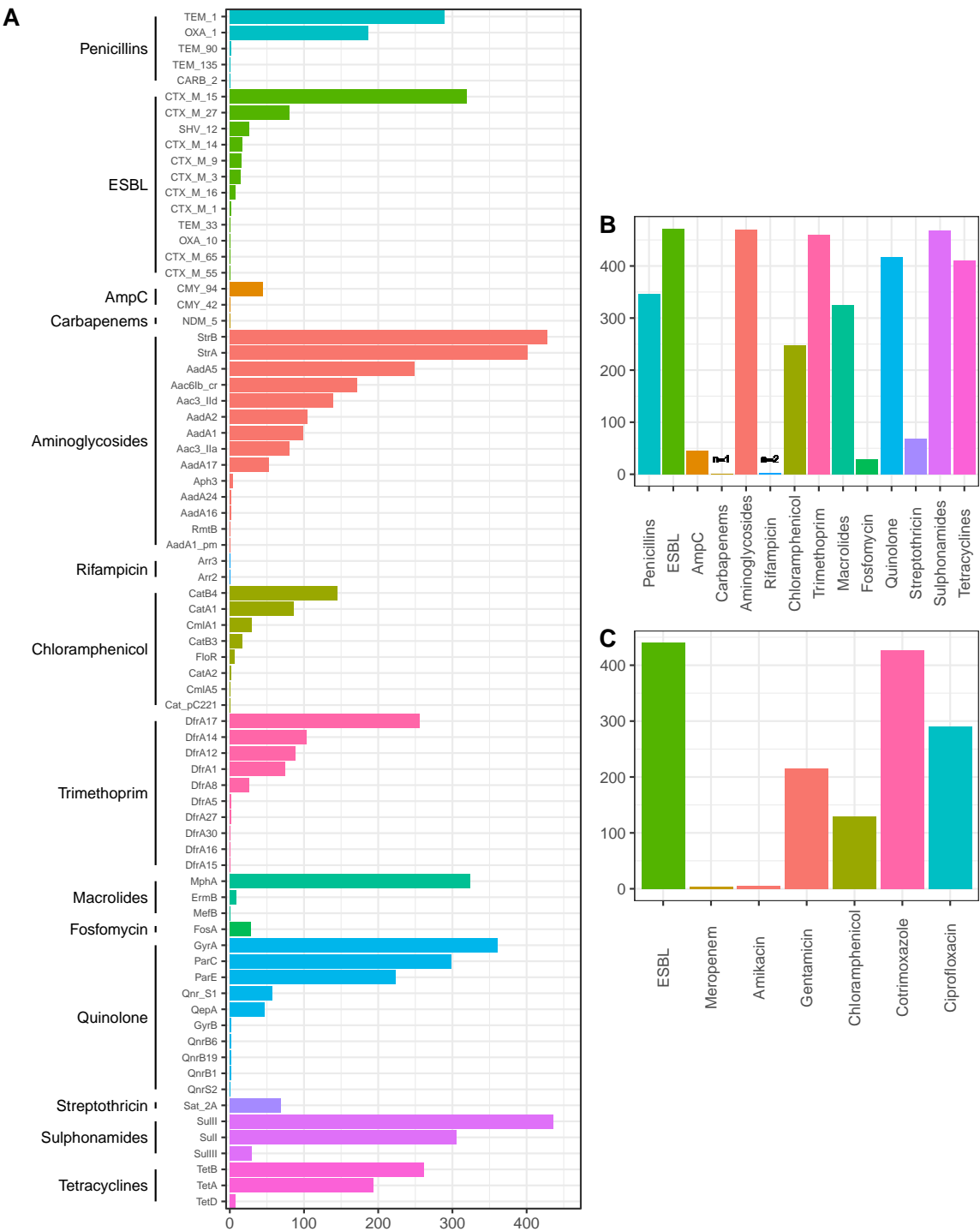


Figure 6.7: A: Frequency distribution of AMR genes identified in isolates. Class of antimicrobial to which gene confers resistance is shown. B: Number of isolates with any mutation to a given class. Any mutation that could possibly confer resistance to a given class is included, including any mutation in the QRDR for quinolones. C: Phenotypic resistance patterns for subset of samples in this analysis that also underwent phenotypic testing ( $n = 449$ )

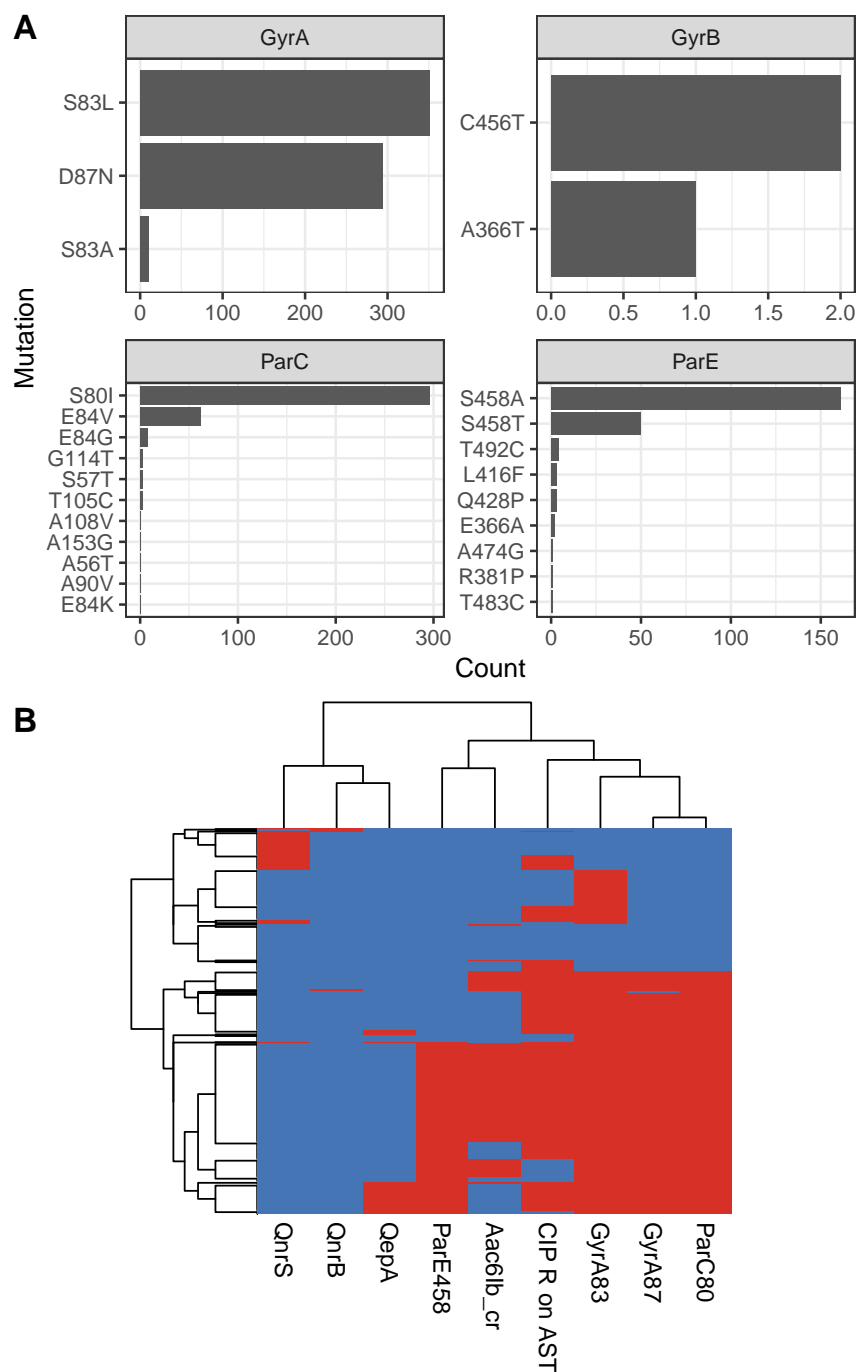


Figure 6.8: A: Mutations in quinolone resistance-determining regions, showing that most mutations are well-recognised (see text for details) B: Row and column clustered heatmap of commonest QRDR mutations (*gyrA83*, *gyrA87*, *parC80* or *parE458*), plasmid-mediated quinolone resistance mutations (*qnr*, *qep* or *aac(6)-Ib-cr*) and phenotypic resistance. The constellation of *gyrA83*, *gyrA87* and *parC80* is strongly associated with phenotypic resistance, but *gyrA83* or *qnrS* alone are insufficient to confer resistance. Each row is one sample, red = presence, blue = absence. CIP R on AST indicates those samples which show phenotypic ciprofloxacin resistance on antimicrobial sensitivity testing (AST). Those isolates without AST data are excluded from this heatmap.

enzyme identifier b. All enzymes of a given class and subclass would be expected to confer similar resistance. There is also a second nomenclature, where genes are referred to *aac*, *aad* (instead of ANT) or *aph*, along with a letter to indicate side of modification; so *aadA*, for example, is equivalent to ANT(3'')-Ia[354]. There are some genes that confer aminoglycoside resistance in different ways and are not included in this classification: the streptomycin resistance genes *strA* and *strB* and the 16S rRNA methylase *rmtB* that confers resistance to amikacin, gentamicin and gentamicin[355].

Aminoglycoside resistance genes were very common in the sequenced isolates, with 469/473 (99%) of isolates containing at least one aminoglycoside resistance gene, and most containing multiple different genes: median number of aminoglycoside resistance genes per isolate was 4 (IQR 3-5). Despite streptomycin being absent from all Malawian treatment guidelines save for retreatment of tuberculosis, the streptomycin resistance genes *strA* and *strB* were near ubiquitous (Figure 6.9). Genes encoding the AAC and ANT enzyme families were commonly seen: genes encoding ANT family enzymes were all *aadA* alleles (alternately known as *aad(3'')-Ia*) which would be expected to also confer streptomycin resistance. Identified genes encoding AAC family enzymes were *aac(3)-Ib* and *aac(3)-IIa*, both of which have been associated with gentamicin resistance, as has *aac(6')-Ib-cr* which also can confer resistance to quinolones. Genes encoding APH family enzymes were unusual. Four were identified, and were all *aph(3')-Ia*, which has been associated with amikacin and kanamycin resistance. One *rmtB* gene was identified, which again has been associated with amikacin resistance.

Genes of the *aac* family tended to co-occur (Figure 6.9B) and presence of any *aac* family gene was strongly associated with gentamicin resistance (OR 9.3 [95% CI 6.0-14.8],  $p < 0.001$ ). Of the five isolates that were resistant to amikacin on AST, two contained *aac* family genes, five contained *aad* family genes and none contained *rmtB* or *aph* genes.

#### 6.4.4.4 Chloramphenicol resistance

Presence of chloramphenicol resistance determinants was common; 248/473 (52%) of isolates contained at least one chloramphenicol resistance gene (Figure 6.7), usually 1 (210/248 [85%]), less commonly 2 (37/248 [15%]) or 3 (1/248 [ $<1\%$ ]). The most commonly identified gene was *catB4*, but presence of *catB4* was not associated with phenotypic chloramphenicol resistance (OR 0.9 [95% CI 0.6-1.4],  $p = 0.65$ ). In comparison, presence of any other chloramphenicol resistance gene was associated with phenotypic resistance (OR 2.5 [95% CI 1.6-3.9],  $p < 0.001$  for a composite variable of all other genes). The reason for this is not clear, but in addition partially assembled *catB4* genes were very common; of the 328 isolates in which there was no fully assembled *catB4*, 93% (306/323) were reported by ARIBA to contain a partially assembled *catB4* gene. In many cases, these partial genes

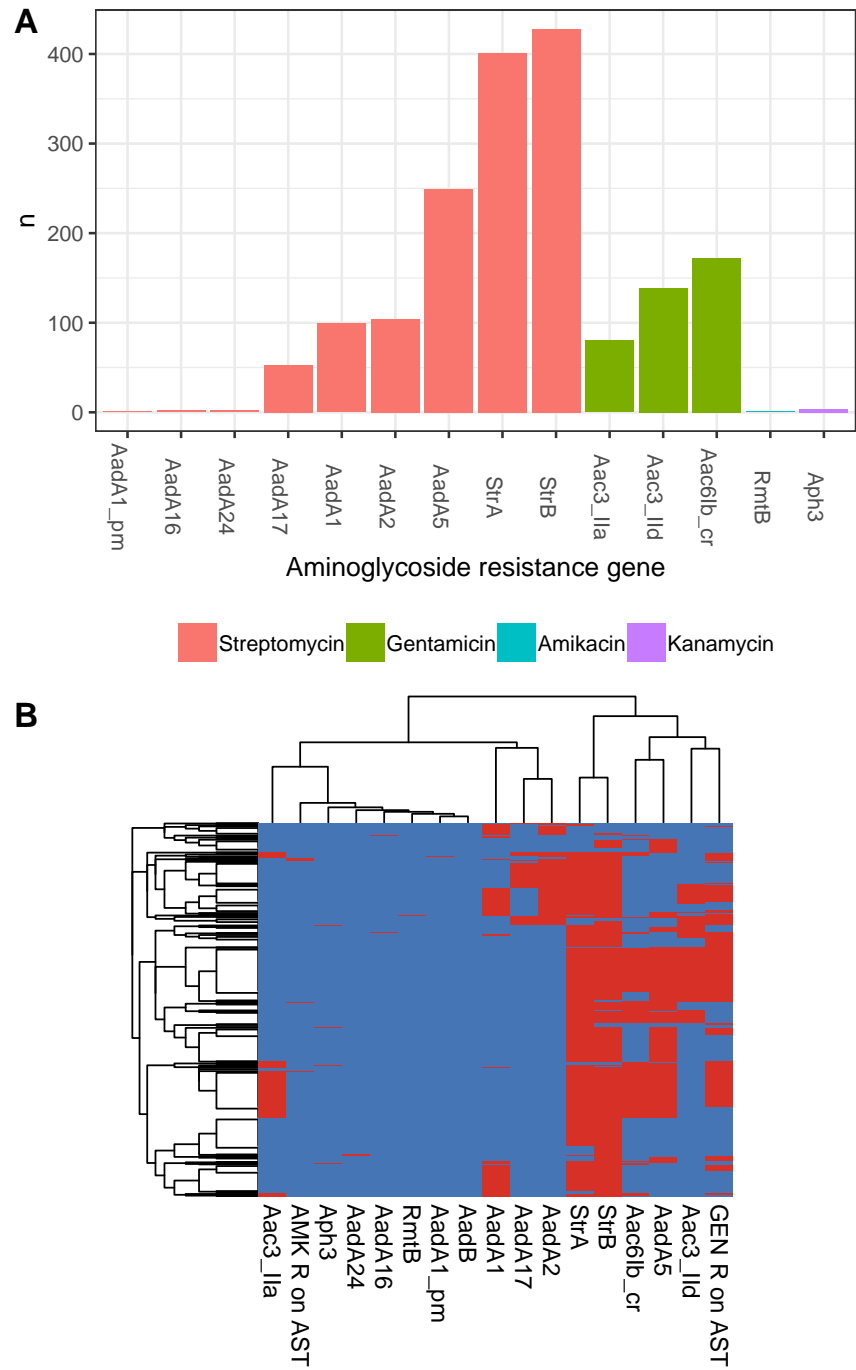


Figure 6.9: A: Aminoglycoside mutations classified by expected resistance to gentamicin, amikacin and kanamycin B: Row and column clustered heatmap showing phenotypic amikacin and gentamicin resistance and identified aminoglycoside resistance genes. AadA indicates *aadA* genes, also referred to as ANT(3'')-Ia. Aac3\_Ila and \_IId refers to *aac(3)-IIa* and *-IId* genes respectively, and Aac6\_Ib\_cr to *aac(6')-Ib-cr*. Aph3 refers to *aph(3')-Ia*. GEN R on AST and AMK R on AST refer to samples that were resistant to gentamicin or amikacin on antimicrobial sensitivity testing (AST). Each row is one sample, red = presence, blue = absence. Samples lacking AST data were excluded from this heatmap.

seemed to be truncated by an IS26 insertion sequence (see Chapter 8).

#### 6.4.4.5 Co-trimoxazole, tetracycline, macrolide and other resistance determinants

Almost all isolates contained either a trimethoprim resistance (459/473 [97%]) or a sulphonamide resistance gene (468/473 [99%]); only 3/473 isolates did not contain either. Trimethoprim resistance genes were all of the *dfrA* family; *sulII* was the commonest sulphonamide resistance determinant (Figures 6.7 and 6.10B). Phenotypic cotrimoxazole resistance was also near-ubiquitous, in 96% (433/448) of those isolates in which antimicrobial sensitivity testing was done.

Tetracycline and macrolide resistance genes were also very common: tetracycline resistance determinants were identified in 411/473 (86%) of isolates, most commonly *tetB* (262/473 [55%] of isolates) and *tetA* (193/473 [41%] of isolates). Macrolide resistance determinants were detected in 325/473 (69%); all isolates carried *mphA* most commonly alone but rarely with *ermB* (9 isolates) or *mefB* (1 isolate). No antimicrobial sensitivity testing was carried out for any agent of the tetracycline or macrolide class. Resistance determinants for rifampicin (*arr2* and *arr3*) were rarely identified, in 2 isolates and the *sat2* gene, conferring resistance to streptothricin (a nucleoside antibiotic with no clinical compounds in use) was seen in 69/473 [15%] of isolates; the significance of this is unknown. Finally, the fosfomycin resistance determinant *fosA* was seen in 28/473 [6%] of isolates, despite this antimicrobial being unavailable in Malawi. It was not restricted to any one clade (Figure 6.13).

#### 6.4.4.6 Clustering and lineage association of AMR determinants

Next, I explored associations of AMR determinants, both with each other in an attempt to identify putative clusters that could represent mobile genetic elements (MGE) that could be tracked within and between patients, and with lineages of the phylogeny. There was clear clustering of AMR genes beyond what would be expected by chance (Figures 6.11 and 6.12), including clustering of the ESBL gene *bla<sub>CTXM-15</sub>* with penicillinases *bla<sub>OXA-1</sub>* and *bla<sub>TEM-1</sub>*. Though some identified clusters correspond to known MGE (e.g. the *sulII-strA-strB* cluster[356]), there was a clear lineage association of certain gene combinations on mapping the presence or absence of AMR determinants back to the phylogeny (Figure 6.12), meaning that these AMR-gene associations likely represent a combination of co-location on MGE and confounding by association with lineage, and suggesting that using clusters of AMR genes to track MGE within participants is likely to be confounded by lineage.



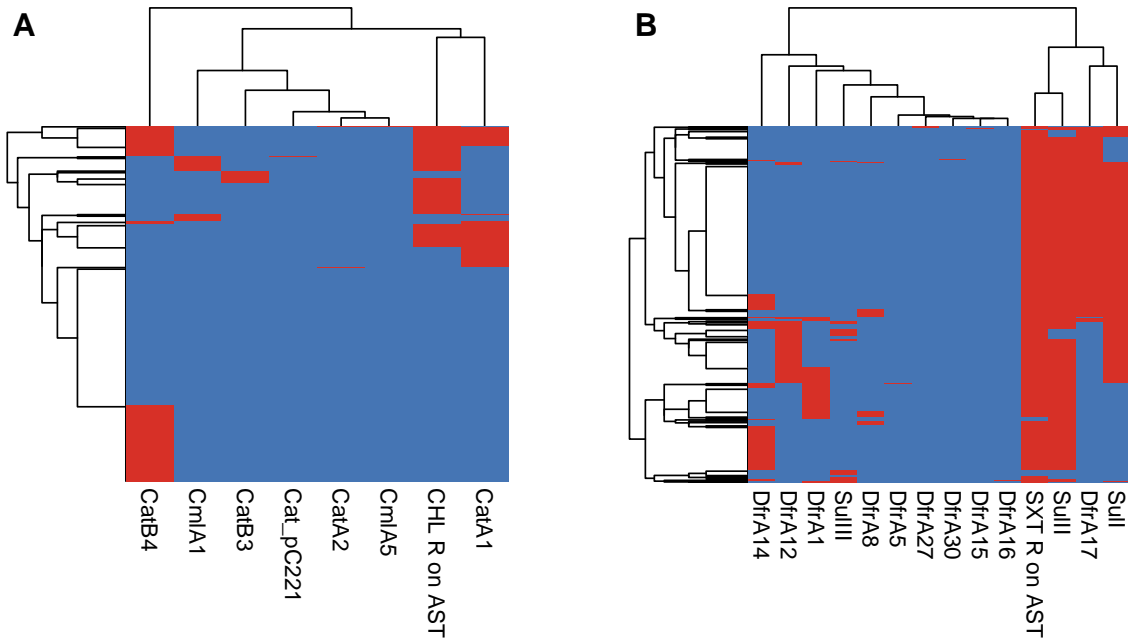


Figure 6.10: Heatmap showing phenotypic chloramphenicol (A) and cotrimoxazole (B) resistance and identified resistance genes that could be expected to confer resistance to these agents. Each row is one sample, red = presence, blue = absence. Both heatmaps are row and column clustered.

### 6.4.5 Plasmid replicons

The frequency of isolation of different plasmid replicons is shown in Table 6.2 and presence or absence of the identified plasmid replicons is shown mapped to the phylogeny in Figure 6.13. IncFIB was most commonly identified (399/473 [84%] of isolates), followed by IncFII (383/473 [81%] of isolates) and IncF1a (324/373 [68%] of isolates). Col plasmids were also frequently identified, in 308/473 [65%] of isolates. Once again, there seems to be some lineage associations of presence or absence of replicons (Figure 6.13).

Table 6.2: Frequency of identification of plasmid replicons

Plasmid replicon	n
IncFIB	399
IncFII	383
IncFIA	324
Col	308
IncY	57
IncI1	40
IncX4	22
IncX1	21
IncHI2	19
IncI2	19
IncR	13
IncB	12
IncN	6
IncHI1	4

Table 6.2: Frequency of identification of plasmid replicons (*continued*)

Plasmid replicon	n
IncL	4
IncA_C	2
IncX2	2
IncFIC	1
IncX3	1

## 6.5 Discussion

In this chapter, I have presented the results of whole genome sequencing of 473 ESBL *E. coli* recovered from serial sampling of 230 Malawian adults from a combination of healthcare-associated and community settings. I have placed these Malawian isolates in the context of the global diversity of *E. coli* using phylogroup, ST and phylogenetic modelling. I have described the AMR determinants and plasmids present in the isolate of and explored clustering of AMR genes.

### 6.5.1 Genomic landscape of ESBL *E. coli* in Malawi: global diversity and high-risk clones

The *E. coli* recovered from stool of the study participants in this study are diverse, encompassing the spectrum of diversity of the species with all major phylogroups and 56 STs represented. Phylogroup A was the commonest phylogroup seen, consistent with the traditional view of this phylogroup as associated with commensal strains[350]. When placed into the context of genomes from throughout the world, the Malawian isolates are largely distributed throughout the phylogeny: in a global context, Malawi is sampling the worldwide diversity of *E. coli*.

There were, however, several areas of the global phylogeny where the Malawian isolates clustered tightly together, perhaps initially suggestive of Malawi-restricted clones; however in considering the significance of this tree topology it is important to be cognisant of the biases inherent in the global *E. coli* collection. ESBL-producing *E. coli* are unusual in the ETEC[347] and GEMS[348] collections and all samples in these two studies were collected before 2011. Both of these collections are exclusively recovered from stool. In contrast, the clinical isolates from the Thai study[346] are all invasive, from a single centre, are selected on the basis of being ESBL-producers, and were isolated in 2014 or 2015. The isolates from the previous Malawian study were largely invasive[279], were selected for diversity in AMR profile, and were all isolated before 2014. There was no study that selectively cultured for ESBL producing *E. coli* in stool, as this study has done; in that, all of these studies are slightly sub-optimal for comparison. It may be that these biases contribute

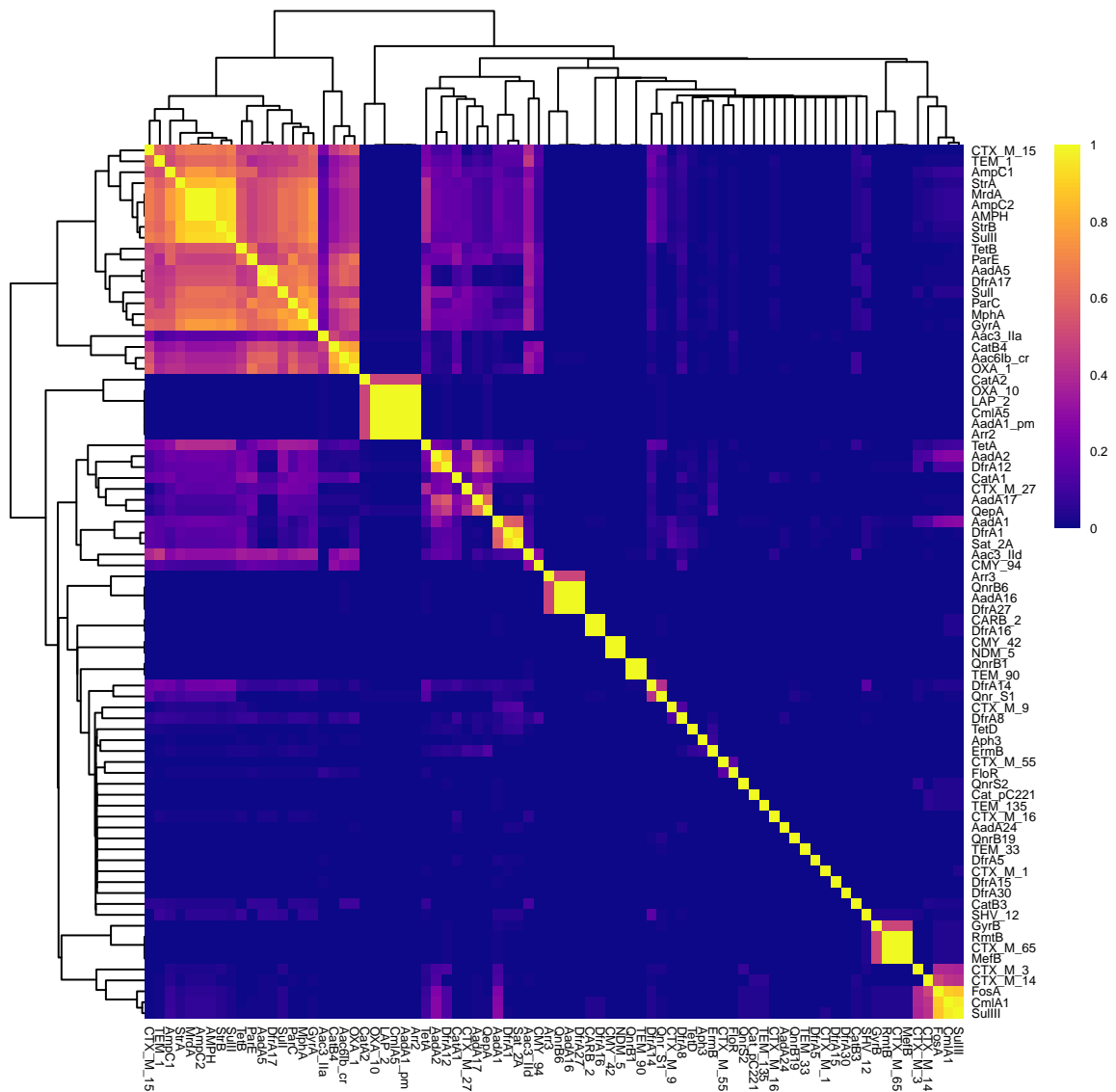


Figure 6.11: Row and column clustered heatmap of pairwise Jaccard index matrix, showing clustering of AMR genes.

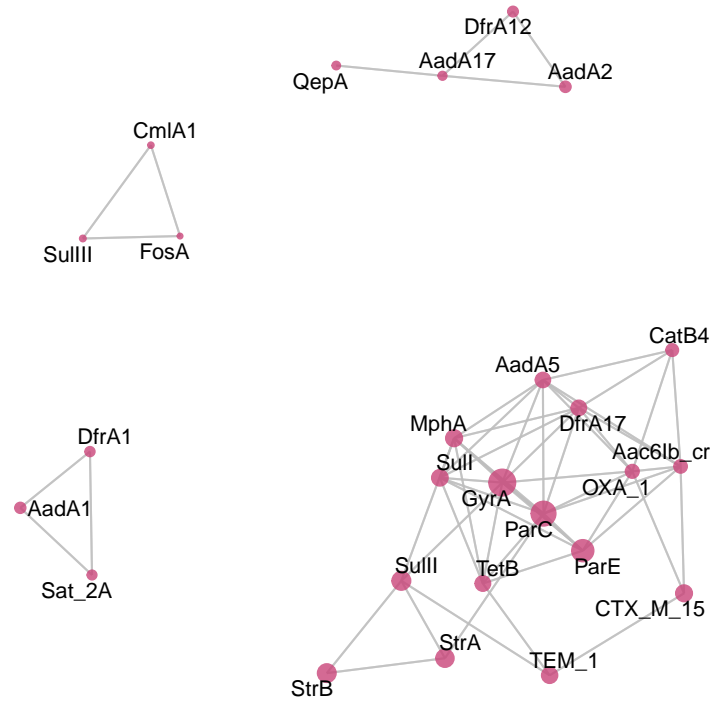


Figure 6.12: Networks of associated (jaccard index > 0.5) AMR genes that occur more often than expected by chance (Bonferroni corrected p-value < 0.05).

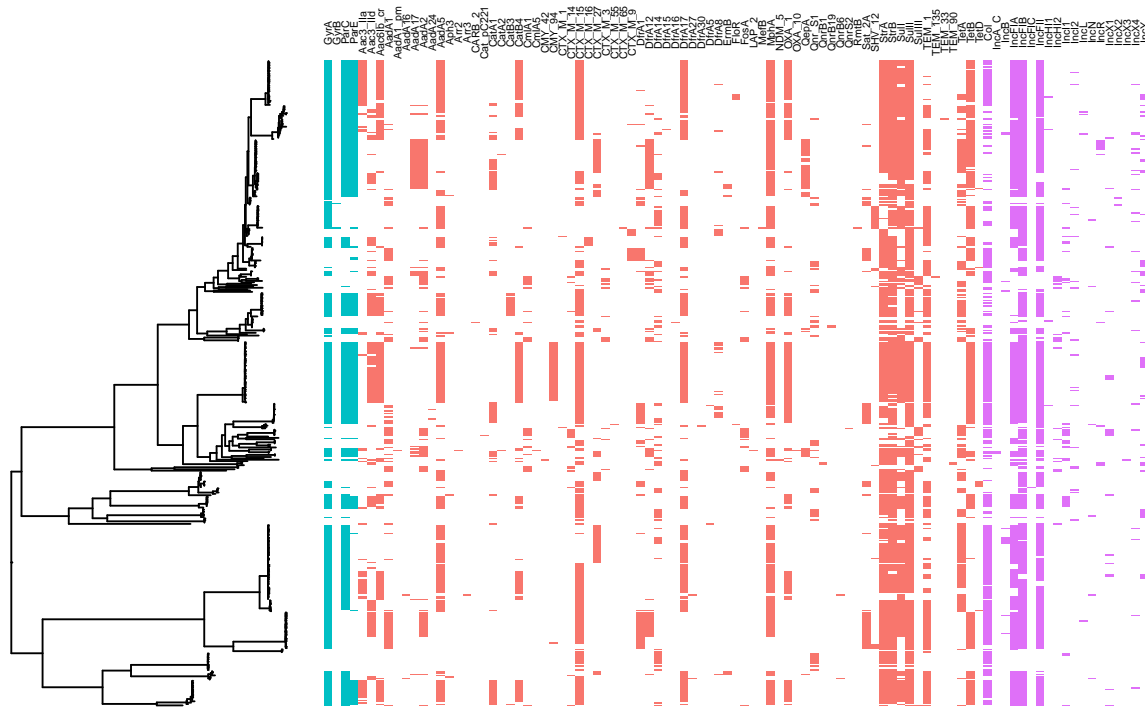


Figure 6.13: AMR genes (blue, chromosomal quinolone resistance and red, other AMR genes) and plasmid replicons (purple) mapped back to tree showing that some AMR gene associations are also associated with lineage.

to apparent polyphyletic clustering of isolates from the current study in phylogroup A in the global tree. It would be expected that ESBL producing phylogroup A *E. coli* would be under-represented in the global collection compared to this study, as this phylogroup is associated with commensal (and hence stool) associated strains, and the two studies performing stool culture did not enrich for ESBL producers; the only study to do this collected invasive isolates, which may be expected to lie in phylogroup B2 over A.

Nevertheless, two of the three commonest STs identified in this study, ST410 and ST167, are unusual or absent in the global collection and could be considered to have a single introduction into Malawi in the context of the topology of the inferred phylogenetic tree. These could represent Malawi-restricted clades or, more likely given the diversity otherwise seen in the tree, clades that are not represented in the global collection because of biases in sample selection. ST410 is recognised as an emerging high-risk clone which has been with isolated worldwide with some regularity since 2011 (including in Tanzania) and is associated with *bla*<sub>CTXM-15</sub> and *bla*<sub>NDM-5</sub>; coalescence analysis suggested a most recent common ancestor of ST410 of the early 1800s (similar to ST131[189]), and acquisition of *bla*<sub>CTXM-15</sub> on a multireplicon IncFII-IncFIA-IncFIB plasmid in the late 1980s[357]. Similarly, ST167 has been recognised as commonly carrying ESBL genes and carbapenemases in Chinese invasive isolates[358] - indeed, it was the commonest *E. coli* ST in one longitudinal surveillance study of carbapenemases in 2012-16 in 25 Chinese provinces[359] - as well as being very prevalent among meat-associated *E. coli* in Germany between 2011-2013 in one study[360]. As such it, too, is also likely a successful global AMR-associated lineage. It is therefore likely that these STs are not represented in the global tree because they have recently emerged worldwide, and have also recently arrived in Malawi. If this is the case, then ST410 and 167 have become rapidly established in Blantyre over the course of only 2-3 years (as they were not seen at all in the previous Malawian strain collection); in fact, this is exactly what was seen in longitudinal nationwide genomic surveillance of *E. coli* in the UK in 2003-04 when ST131 first arrived[361].

ST131 is a globally established high risk clone and the commonest ST in this study and the global collection. Indeed this ST is thought to account for 40-80% of invasive ESBL *E. coli* infection worldwide[188,362]. The topography of the global tree suggests no obvious barriers to mixing between Malawian and global ST131. It may be that unbiased global sampling would reveal the same pattern for ST167 and ST410. Though some progress has been made in understanding the genomics of the emergence of ST131[189], the factors that contribute to its apparent fitness are unknown: it is impossible to predict, at present, from the genome of ST167 and ST410 whether they will repeat the course of ST131 to become truly globally dominant as a cause of human disease. Such an understanding of the determinants of fitness would be of great benefit in predicting and preventing global AMR spread.

### 6.5.2 Antimicrobial resistance determinants: domination of *bla*<sub>CTXM-15</sub> and emergence of carbapenemases

The 473 isolates contained a diverse selection of antimicrobial resistance determinants, most with genotypic multiclass resistance. Genotypic and phenotypic co-trimoxazole resistance was near universal, as might be expected from a setting where lifelong co-trimoxazole preventative therapy (CPT) is mandated by the Malawian HIV treatment guidelines for HIV all infected adults[330], and mediated by *dfrA* and *Sul* alleles.

Determinants of aminoglycoside resistance were also present in almost every sample, most commonly streptomycin resistance determinants, despite the lack of drug pressure. Well recognised gentamicin-resistance determinants - *aac(3)* and *aac(6'')* - were strongly associated with gentamicin resistance, but the aminoglycoside resistance genes present in isolates displaying apparent amikacin resistance (*aad5* and *aac(3)*) would not usually be expected to confer amikacin resistance. Given the extreme rarity of phenotypic amikacin resistance, re-testing these isolates with a more accurate AST method should be the first step (see limitations, below).

Quinolone resistance determinants were also common. Quinolones are widely used in Blantyre, and are the current treatment of choice for invasive *Salmonella* infections, one of the commonest cause of bloodstream infection here[15]. The genotypic determinants of quinolone resistance are complex, mediated by point mutations in the drug target regions, acquisition of modifying enzymes or up regulation of multidrug resistance pumps, and usually multiple genes or mutations are needed to bring about a resistant phenotype[363]. This is the pattern observed in this dataset where the proportion of isolates with identifiable determinants of quinolone resistance was greater than those with phenotypic resistance. Nevertheless, the genotype-phenotype associations seen in this data are largely those that have been described in the literature: the *gyrA83-parC80-gyrA87* mutation combination has been shown to strongly predict combination quinolone resistance in a study of 10099 *E. coli* genomes[364], and it is recognised that the presence of the *qnrS* gene alone is usually not sufficient to bring about a resistant phenotype[365].

Presence of chloramphenicol resistance genes, particularly *catB4*, was more common than phenotypic chloramphenicol resistance such that *catB4* was not associated with phenotypic resistance. This is unexpected, as presence of chloramphenicol resistance genes has been shown to correlate well with phenotypic resistance[366], including in a study of 94 Malawian invasive isolates[367], though in this Malawian collection (the same collection of isolates as were included in the global collection in the current study), *catB* genes were rarely seen. Interestingly, truncated *catB4* elements (often in conjunction with an IS26 transposon) were almost universal in the isolates in this study: of 323 isolates in which ARIBA did not assemble a full *catB4* sequence, 93% (306/323) contained a truncated *catB4*

element. This configuration (*catB4* truncated by an IS26 element) has been described in Enterobacteriaceae[368,369]. It could certainly be unrelated but its ubiquity in this study raises at least the possibility of misassembly and false-positive identification of *catB4* in some cases. Laboratory errors resulting in erroneous AST should also be considered. It is also possible that the *catB4* gene in this collection is not expressed or downregulated in some way; exploring these hypotheses is a possible direction for future work.

ESBL resistance in this collection is dominated by *bla<sub>CTXM</sub>* and *bla<sub>CTXM-15</sub>* in particular; this latter gene was carried by 319/473 [67%] of isolates. The only non-*bla<sub>CTXM</sub>* ESBL gene identified in any significant prevalence was *bla<sub>SHV-12</sub>*; ESBL *bla<sub>OXA</sub>* and *bla<sub>TEM</sub>* were very rare, though narrow-spectrum penicillinase alleles of this family were common. The dominance of *bla<sub>CTXM-15</sub>* is in keeping with the situation seen worldwide[155]. In this collection, *bla<sub>CTXM-15</sub>* was spread throughout the phylogeny rather than associated with any particular clade, as would be expected from its global dominance. I identified one carbapenemase, *bla<sub>NDM-5</sub>*, carried on a globally successful IncX3 plasmid. To my knowledge, this is the first carbapenemase to be described in Malawi. Carbapenem antimicrobials were introduced to the Malawian essential medicines list in 2015 but are at best sporadically available, and only in tertiary centres, often for truncated courses. The emergence of carbapenemases with such minimal carbapenem use and so soon after introduction is troubling, and should prompt discussions regarding the best use of this precious antimicrobial class; certainly, given the high prevalence of ESBL production among invasive *K. pneumoniae* and *E. coli*[15], there is a case for expanded access but optimal antimicrobial stewardship strategies in this context are unknown.

### 6.5.3 Study limitations

There are several limitations to the analysis carried out in this chapter. Only one *E. coli* pick from each time point was sequenced which may have missed significant within-participant diversity. This is seriously problematic for the analyses of the next chapter; here, diversity may have been missed but the data presented here should at least reflect a random sample from the *E. coli* carried by the participants in the study. Community isolates are under represented, largely because of a lower ESBL-E prevalence in community members. The sequenced *E. coli* isolates presented here do not represent all the samples in the study from which ESBL producing *E. coli* were isolated, but only those up to the point at which the samples were shipped for sequencing. I have focused on *E. coli*; in fact, these were often isolated in conjunction with other bacteria, most notably *Klebsiella pneumoniae*, which were not induced in this analysis. Given the propensity of *K. pneumoniae* to carry AMR genes, the circulating AMR gene diversity in Enterobacteriaceae carried by participants in this study is likely greater than I describe here.

For context genomes, I selected a global collection of *E. coli* based on what was available but, as described above and in common with many analyses of this type, this is a biased collection. This must be borne in mind when interpreting the global phylogeny. There are inherent limitations in the short-read Illumina sequencing that was carried out: assembly of areas with multiple nucleotide repeats (as found in plasmids and transposable elements in particular) is difficult or impossible, making it impossible to fully characterise the MGE in this dataset upon which the AMR genes are carried.

There are some discrepancies between phenotypic resistance and what would be expected from identified genes. Partly phenotype may have been misclassified. The AST method used was disc diffusion; certainly this is less accurate than an MIC determination method such as E-tests or dilution methods, and a true comparison of genotypic to phenotypic resistance - not the aim of this study - should use one of these methods. It is also possible that there were technical problems with the AST procedure (e.g. an overly heavy inoculum) though every attempt was made to avoid this, including with internal QC, and the work was carried out in a laboratory which subscribes to the UK NEQAS QC procedure.

#### 6.5.4 Conclusions and further work

In conclusion, I have shown that the *E. coli* population in this study is diverse, representing global *E. coli* diversity, and suggesting significant mixing of *E. coli* between Blantyre and the rest of the world. The AMR genes identified seem to reflect the local antibiotic pressures: near universal cotrimoxazole resistance, moderate quinolone resistance and very little carbapenem resistance. The presence of carbapenem resistance mediated by a *bla<sub>NDM-5</sub>* carbapenemase is to my knowledge the first carbapenemase described in Malawi. Its presence along with apparently unrestricted mixing of *E. coli* from Malawi with the rest of the world is a reminder that international spread of bacteria and AMR determinants, even to low-resource settings like Malawi, can be rapid. Treatment options for invasive Gram-negative infections in Malawi are often limited by ESBL resistance[15]. There is a significant unmet need for carbapenem antibiotics, but antibiotic stewardship interventions that can balance access to against restriction of these valuable last-line antimicrobials are needed to prevent rapid emergence of resistance following wider-scale carbapenem roll-out in Malawi.

Further work is planned:

- The discordance between chloramphenicol phenotype and genotype should be explored, especially as this antibiotic - largely fallen from favour in Malawi since ceftriaxone has been available - may have a role to play in treating some invasive ESBL-E infections. Long-read sequencing of representative isolates will exclude the possibility of missassembly and allow assessment of the genetic environment of the *catB4* gene;



it is possible that it has been downregulated or is not expressed, for example. More accurate chloramphenicol AST using e.g. E-tests may also be helpful to confirm phenotype, which should also be done with the apparent amikacin resistance isolates presented here.

- The apparent recent arrival of the putative high-risk clones ST410 and ST167 deserves further attention. The isolates here can be placed in a global context with other publicly available ST410 and ST167 *E. coli*, but a question remains as to whether these clones are causing invasive disease. This question can only be answered by sequencing invasive isolates; this work is ongoing and will allow such a comparison to be made.
- The remainder of the samples (post October 2018) will have one *E. coli* sent for sequencing.
- All samples where *K. pneumoniae* were isolated will have one *K. pneumoniae* colony pick sequenced allowing a comparative analysis of AMR determinants by *K. pneumoniae* and *E. coli* carried within a single participant.
- More broadly, I have presented a description of the isolates sequenced in this study, but largely without reference to the available metadata. The motivation in sequencing these isolates was to use WGS as a high-resolution typing scheme to track bacteria within and between participants. This analysis is the focus of the next chapter.

## 6.6 Appendix

Table 6.3: Details of samples included in global phylogeny

Study	Sample ID	Accession No.	Source	Country
This study	26141_1_134	ERS2493389	Stool	Malawi
This study	26141_1_135	ERS2493390	Stool	Malawi
This study	26141_1_136	ERS2493391	Stool	Malawi
This study	26141_1_137	ERS2493392	Stool	Malawi
This study	26141_1_138	ERS2493393	Stool	Malawi
This study	26141_1_139	ERS2493394	Stool	Malawi
This study	26141_1_140	ERS2493395	Stool	Malawi
This study	26141_1_141	ERS2493396	Stool	Malawi
This study	26141_1_142	ERS2493397	Stool	Malawi
This study	26141_1_143	ERS2493398	Stool	Malawi
This study	26141_1_144	ERS2493399	Stool	Malawi
This study	26141_1_145	ERS2493400	Stool	Malawi
This study	26141_1_146	ERS2493401	Stool	Malawi
This study	26141_1_147	ERS2493402	Stool	Malawi
This study	26141_1_148	ERS2493403	Stool	Malawi
This study	26141_1_149	ERS2493404	Stool	Malawi
This study	26141_1_150	ERS2493405	Stool	Malawi
This study	26141_1_151	ERS2493406	Stool	Malawi
This study	26141_1_152	ERS2493407	Stool	Malawi
This study	26141_1_153	ERS2493408	Stool	Malawi
This study	26141_1_154	ERS2493409	Stool	Malawi

Table 6.3: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
This study	26141_1_155	ERS2493410	Stool	Malawi
This study	26141_1_156	ERS2493411	Stool	Malawi
This study	26141_1_157	ERS2493412	Stool	Malawi
This study	26141_1_158	ERS2493413	Stool	Malawi
This study	26141_1_159	ERS2493414	Stool	Malawi
This study	26141_1_160	ERS2493415	Stool	Malawi
This study	26141_1_161	ERS2493416	Stool	Malawi
This study	26141_1_162	ERS2493417	Stool	Malawi
This study	26141_1_164	ERS2493419	Stool	Malawi
This study	26141_1_165	ERS2493420	Stool	Malawi
This study	26141_1_166	ERS2493421	Stool	Malawi
This study	26141_1_168	ERS2493423	Stool	Malawi
This study	26141_1_169	ERS2493424	Stool	Malawi
This study	26141_1_170	ERS2493425	Stool	Malawi
This study	26141_1_171	ERS2493426	Stool	Malawi
This study	26141_1_172	ERS2493427	Stool	Malawi
This study	26141_1_173	ERS2493428	Stool	Malawi
This study	26141_1_174	ERS2493429	Stool	Malawi
This study	26141_1_175	ERS2493430	Stool	Malawi
This study	26141_1_176	ERS2493431	Stool	Malawi
This study	26141_1_177	ERS2493432	Stool	Malawi
This study	26141_1_178	ERS2493433	Stool	Malawi
This study	26141_1_179	ERS2493434	Stool	Malawi
This study	26141_1_180	ERS2493435	Stool	Malawi
This study	26141_1_181	ERS2493436	Stool	Malawi
This study	26141_1_182	ERS2493437	Stool	Malawi
This study	26141_1_183	ERS2493438	Stool	Malawi
This study	26141_1_184	ERS2493439	Stool	Malawi
This study	26141_1_186	ERS2493441	Stool	Malawi
This study	26141_1_187	ERS2493442	Stool	Malawi
This study	26141_1_189	ERS2493444	Stool	Malawi
This study	26141_1_190	ERS2493445	Stool	Malawi
This study	26141_1_191	ERS2493446	Stool	Malawi
This study	26141_1_192	ERS2493447	Stool	Malawi
This study	26141_1_193	ERS2493448	Stool	Malawi
This study	26141_1_194	ERS2493449	Stool	Malawi
This study	26141_1_195	ERS2493450	Stool	Malawi
This study	26141_1_196	ERS2493451	Stool	Malawi
This study	26141_1_197	ERS2493452	Stool	Malawi
This study	26141_1_198	ERS2493453	Stool	Malawi
This study	26141_1_199	ERS2493454	Stool	Malawi
This study	26141_1_200	ERS2493456	Stool	Malawi
This study	26141_1_201	ERS2493457	Stool	Malawi
This study	26141_1_202	ERS2493458	Stool	Malawi
This study	26141_1_203	ERS2493459	Stool	Malawi
This study	26141_1_204	ERS2493460	Stool	Malawi
This study	26141_1_205	ERS2493461	Stool	Malawi
This study	26141_1_206	ERS2493463	Stool	Malawi
This study	26141_1_207	ERS2493464	Stool	Malawi
This study	26141_1_208	ERS2493465	Stool	Malawi
This study	26141_1_209	ERS2493466	Stool	Malawi
This study	26141_1_210	ERS2493468	Stool	Malawi
This study	26141_1_211	ERS2493469	Stool	Malawi
This study	26141_1_212	ERS2493470	Stool	Malawi
This study	26141_1_213	ERS2493471	Stool	Malawi
This study	26141_1_214	ERS2493472	Stool	Malawi
This study	26141_1_215	ERS2493473	Stool	Malawi
This study	26141_1_217	ERS2493475	Stool	Malawi
This study	26141_1_218	ERS2493476	Stool	Malawi
This study	26141_1_219	ERS2493477	Stool	Malawi
This study	26141_1_220	ERS2493478	Stool	Malawi
This study	26141_1_221	ERS2493479	Stool	Malawi

Table 6.3: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
This study	26141_1_222	ERS2493480	Stool	Malawi
This study	26141_1_223	ERS2493481	Stool	Malawi
This study	26141_1_224	ERS2493483	Stool	Malawi
This study	26141_1_225	ERS2493484	Stool	Malawi
This study	26141_1_226	ERS2493485	Stool	Malawi
This study	26141_1_227	ERS2493486	Stool	Malawi
This study	26141_1_228	ERS2493487	Stool	Malawi
This study	26141_1_229	ERS2493488	Stool	Malawi
This study	26141_1_230	ERS2493489	Stool	Malawi
This study	26141_1_232	ERS2493491	Stool	Malawi
This study	26141_1_236	ERS2493508	Stool	Malawi
This study	26141_1_237	ERS2493509	Stool	Malawi
This study	26141_1_239	ERS2493511	Stool	Malawi
This study	26141_1_240	ERS2493512	Stool	Malawi
This study	26141_1_241	ERS2493513	Stool	Malawi
This study	26141_1_242	ERS2493515	Stool	Malawi
This study	26141_1_243	ERS2493516	Stool	Malawi
This study	26141_1_244	ERS2493517	Stool	Malawi
This study	26141_1_246	ERS2493519	Stool	Malawi
This study	26141_1_247	ERS2493520	Stool	Malawi
This study	26141_1_248	ERS2493521	Stool	Malawi
This study	26141_1_250	ERS2493495	Stool	Malawi
This study	26141_1_251	ERS2493496	Stool	Malawi
This study	26141_1_252	ERS2493497	Stool	Malawi
This study	26141_1_253	ERS2493498	Stool	Malawi
This study	26141_1_254	ERS2493499	Stool	Malawi
This study	26141_1_255	ERS2493501	Stool	Malawi
This study	26141_1_256	ERS2493502	Stool	Malawi
This study	26141_1_257	ERS2493503	Stool	Malawi
This study	26141_1_258	ERS2493504	Stool	Malawi
This study	26141_1_259	ERS2493505	Stool	Malawi
This study	26141_1_260	ERS2493506	Stool	Malawi
This study	26141_1_261	ERS2493522	Stool	Malawi
This study	26141_1_262	ERS2493523	Stool	Malawi
This study	26141_1_263	ERS2493524	Stool	Malawi
This study	26141_1_265	ERS2493526	Stool	Malawi
This study	26141_1_266	ERS2493527	Stool	Malawi
This study	26141_1_267	ERS2493528	Stool	Malawi
This study	26141_1_268	ERS2493529	Stool	Malawi
This study	26141_1_270	ERS2493531	Stool	Malawi
This study	26141_1_271	ERS2493532	Stool	Malawi
This study	26141_1_272	ERS2493533	Stool	Malawi
This study	26141_1_273	ERS2493534	Stool	Malawi
This study	26141_1_274	ERS2493537	Stool	Malawi
This study	26141_1_275	ERS2493538	Stool	Malawi
This study	26141_1_276	ERS2493539	Stool	Malawi
This study	26141_1_277	ERS2493540	Stool	Malawi
This study	26141_1_278	ERS2493541	Stool	Malawi
This study	26141_1_279	ERS2493542	Stool	Malawi
This study	26141_1_280	ERS2493543	Stool	Malawi
This study	26141_1_282	ERS2493545	Stool	Malawi
This study	26141_1_283	ERS2493546	Stool	Malawi
This study	26141_1_284	ERR3168700	Stool	Malawi
This study	26141_1_285	ERS2493548	Stool	Malawi
This study	26141_1_286	ERS2493549	Stool	Malawi
This study	26141_1_287	ERS2493550	Stool	Malawi
This study	26141_1_288	ERS2493551	Stool	Malawi
This study	26141_1_289	ERS2493552	Stool	Malawi
This study	26141_1_290	ERS2493553	Stool	Malawi
This study	26141_1_291	ERS2493554	Stool	Malawi
This study	26141_1_292	ERS2493555	Stool	Malawi
This study	26141_1_293	ERS2493556	Stool	Malawi

Table 6.3: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
This study	26141_1_295	ERS2493558	Stool	Malawi
This study	26141_1_296	ERS2493560	Stool	Malawi
This study	26141_1_297	ERS2493561	Stool	Malawi
This study	26141_1_298	ERS2493563	Stool	Malawi
This study	26141_1_299	ERS2493564	Stool	Malawi
This study	28099_1_1	ERS2855095	Stool	Malawi
This study	28099_1_10	ERS2855097	Stool	Malawi
This study	28099_1_100	ERS2855215	Stool	Malawi
This study	28099_1_102	ERS2855046	Stool	Malawi
This study	28099_1_103	ERS2855121	Stool	Malawi
This study	28099_1_104	ERS2855216	Stool	Malawi
This study	28099_1_106	ERS2855047	Stool	Malawi
This study	28099_1_107	ERS2855122	Stool	Malawi
This study	28099_1_11	ERS2855194	Stool	Malawi
This study	28099_1_110	ERS2855048	Stool	Malawi
This study	28099_1_111	ERS2855123	Stool	Malawi
This study	28099_1_112	ERS2855218	Stool	Malawi
This study	28099_1_114	ERS2855049	Stool	Malawi
This study	28099_1_115	ERS2855124	Stool	Malawi
This study	28099_1_116	ERS2855219	Stool	Malawi
This study	28099_1_118	ERS2855050	Stool	Malawi
This study	28099_1_119	ERS2855125	Stool	Malawi
This study	28099_1_120	ERS2855220	Stool	Malawi
This study	28099_1_123	ERS2855126	Stool	Malawi
This study	28099_1_125	ERS2855222	Stool	Malawi
This study	28099_1_127	ERS2855019	Stool	Malawi
This study	28099_1_128	ERS2855127	Stool	Malawi
This study	28099_1_129	ERS2855223	Stool	Malawi
This study	28099_1_131	ERS2855020	Stool	Malawi
This study	28099_1_132	ERS2855128	Stool	Malawi
This study	28099_1_133	ERS2855225	Stool	Malawi
This study	28099_1_135	ERS2855021	Stool	Malawi
This study	28099_1_136	ERS2855129	Stool	Malawi
This study	28099_1_137	ERS2855226	Stool	Malawi
This study	28099_1_139	ERS2855022	Stool	Malawi
This study	28099_1_14	ERS2855098	Stool	Malawi
This study	28099_1_141	ERS2855227	Stool	Malawi
This study	28099_1_143	ERS2855023	Stool	Malawi
This study	28099_1_144	ERS2855131	Stool	Malawi
This study	28099_1_145	ERS2855228	Stool	Malawi
This study	28099_1_148	ERS2855132	Stool	Malawi
This study	28099_1_149	ERS2855229	Stool	Malawi
This study	28099_1_151	ERS2855025	Stool	Malawi
This study	28099_1_152	ERS2855133	Stool	Malawi
This study	28099_1_153	ERS2855230	Stool	Malawi
This study	28099_1_155	ERS2855026	Stool	Malawi
This study	28099_1_156	ERS2855134	Stool	Malawi
This study	28099_1_157	ERS2855231	Stool	Malawi
This study	28099_1_159	ERS2855027	Stool	Malawi
This study	28099_1_160	ERS2855135	Stool	Malawi
This study	28099_1_161	ERS2855232	Stool	Malawi
This study	28099_1_163	ERS2855028	Stool	Malawi
This study	28099_1_165	ERS2855233	Stool	Malawi
This study	28099_1_167	ERS2855029	Stool	Malawi
This study	28099_1_168	ERS2855137	Stool	Malawi
This study	28099_1_169	ERS2855234	Stool	Malawi
This study	28099_1_171	ERS2855030	Stool	Malawi
This study	28099_1_172	ERS2855138	Stool	Malawi
This study	28099_1_173	ERS2855235	Stool	Malawi
This study	28099_1_175	ERS2855031	Stool	Malawi
This study	28099_1_176	ERS2855139	Stool	Malawi
This study	28099_1_177	ERS2855236	Stool	Malawi

Table 6.3: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
This study	28099_1_179	ERS2855032	Stool	Malawi
This study	28099_1_18	ERS2855099	Stool	Malawi
This study	28099_1_180	ERS2855140	Stool	Malawi
This study	28099_1_181	ERS2855237	Stool	Malawi
This study	28099_1_185	ERS2855238	Stool	Malawi
This study	28099_1_187	ERS2855034	Stool	Malawi
This study	28099_1_188	ERS2855141	Stool	Malawi
This study	28099_1_189	ERS2855239	Stool	Malawi
This study	28099_1_19	ERS2855195	Stool	Malawi
This study	28099_1_191	ERS2855035	Stool	Malawi
This study	28099_1_192	ERS2855142	Stool	Malawi
This study	28099_1_193	ERS2855240	Stool	Malawi
This study	28099_1_195	ERS2855036	Stool	Malawi
This study	28099_1_196	ERS2855143	Stool	Malawi
This study	28099_1_199	ERS2855037	Stool	Malawi
This study	28099_1_2	ERS2855192	Stool	Malawi
This study	28099_1_200	ERS2855144	Stool	Malawi
This study	28099_1_203	ERS2855038	Stool	Malawi
This study	28099_1_204	ERS2855145	Stool	Malawi
This study	28099_1_205	ERS2855243	Stool	Malawi
This study	28099_1_207	ERS2855039	Stool	Malawi
This study	28099_1_208	ERS2855146	Stool	Malawi
This study	28099_1_209	ERS2855244	Stool	Malawi
This study	28099_1_211	ERS2855051	Stool	Malawi
This study	28099_1_212	ERS2855147	Stool	Malawi
This study	28099_1_213	ERS2855148	Stool	Malawi
This study	28099_1_214	ERS2855245	Stool	Malawi
This study	28099_1_216	ERS2855052	Stool	Malawi
This study	28099_1_217	ERS2855149	Stool	Malawi
This study	28099_1_218	ERS2855246	Stool	Malawi
This study	28099_1_22	ERS2855100	Stool	Malawi
This study	28099_1_220	ERS2855053	Stool	Malawi
This study	28099_1_221	ERS2855150	Stool	Malawi
This study	28099_1_222	ERS2855247	Stool	Malawi
This study	28099_1_224	ERS2855054	Stool	Malawi
This study	28099_1_225	ERS2855151	Stool	Malawi
This study	28099_1_226	ERS2855248	Stool	Malawi
This study	28099_1_228	ERS2855055	Stool	Malawi
This study	28099_1_229	ERS2855152	Stool	Malawi
This study	28099_1_23	ERS2855196	Stool	Malawi
This study	28099_1_230	ERS2855249	Stool	Malawi
This study	28099_1_232	ERS2855056	Stool	Malawi
This study	28099_1_233	ERS2855153	Stool	Malawi
This study	28099_1_234	ERS2855250	Stool	Malawi
This study	28099_1_236	ERS2855057	Stool	Malawi
This study	28099_1_237	ERS2855154	Stool	Malawi
This study	28099_1_238	ERS2855251	Stool	Malawi
This study	28099_1_240	ERS2855058	Stool	Malawi
This study	28099_1_241	ERS2855155	Stool	Malawi
This study	28099_1_242	ERS2855252	Stool	Malawi
This study	28099_1_244	ERS2855059	Stool	Malawi
This study	28099_1_245	ERS2855156	Stool	Malawi
This study	28099_1_246	ERS2855253	Stool	Malawi
This study	28099_1_248	ERS2855060	Stool	Malawi
This study	28099_1_249	ERS2855157	Stool	Malawi
This study	28099_1_250	ERS2855254	Stool	Malawi
This study	28099_1_252	ERS2855061	Stool	Malawi
This study	28099_1_253	ERS2855158	Stool	Malawi
This study	28099_1_254	ERS2855255	Stool	Malawi
This study	28099_1_256	ERS2855062	Stool	Malawi
This study	28099_1_257	ERS2855160	Stool	Malawi
This study	28099_1_258	ERS2855256	Stool	Malawi

Table 6.3: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
This study	28099_1_26	ERS2855101	Stool	Malawi
This study	28099_1_260	ERS2855063	Stool	Malawi
This study	28099_1_261	ERS2855161	Stool	Malawi
This study	28099_1_264	ERS2855064	Stool	Malawi
This study	28099_1_266	ERS2855258	Stool	Malawi
This study	28099_1_268	ERS2855065	Stool	Malawi
This study	28099_1_269	ERS2855163	Stool	Malawi
This study	28099_1_27	ERS2855197	Stool	Malawi
This study	28099_1_270	ERS2855259	Stool	Malawi
This study	28099_1_272	ERS2855066	Stool	Malawi
This study	28099_1_273	ERS2855164	Stool	Malawi
This study	28099_1_274	ERS2855260	Stool	Malawi
This study	28099_1_277	ERS2855165	Stool	Malawi
This study	28099_1_278	ERS2855269	Stool	Malawi
This study	28099_1_280	ERS2855068	Stool	Malawi
This study	28099_1_281	ERS2855166	Stool	Malawi
This study	28099_1_282	ERS2855270	Stool	Malawi
This study	28099_1_284	ERS2855069	Stool	Malawi
This study	28099_1_285	ERS2855167	Stool	Malawi
This study	28099_1_286	ERS2855271	Stool	Malawi
This study	28099_1_288	ERS2855070	Stool	Malawi
This study	28099_1_289	ERS2855168	Stool	Malawi
This study	28099_1_293	ERS2855169	Stool	Malawi
This study	28099_1_294	ERS2855273	Stool	Malawi
This study	28099_1_297	ERS2855170	Stool	Malawi
This study	28099_1_30	ERS2855102	Stool	Malawi
This study	28099_1_300	ERS2855073	Stool	Malawi
This study	28099_1_301	ERS2855074	Stool	Malawi
This study	28099_1_302	ERS2855171	Stool	Malawi
This study	28099_1_303	ERS2855275	Stool	Malawi
This study	28099_1_305	ERS2855075	Stool	Malawi
This study	28099_1_306	ERS2855172	Stool	Malawi
This study	28099_1_307	ERS2855276	Stool	Malawi
This study	28099_1_309	ERS2855076	Stool	Malawi
This study	28099_1_31	ERS2855198	Stool	Malawi
This study	28099_1_311	ERS2855277	Stool	Malawi
This study	28099_1_313	ERS2855077	Stool	Malawi
This study	28099_1_314	ERS2855174	Stool	Malawi
This study	28099_1_315	ERS2855278	Stool	Malawi
This study	28099_1_317	ERS2855078	Stool	Malawi
This study	28099_1_318	ERS2855175	Stool	Malawi
This study	28099_1_319	ERS2855279	Stool	Malawi
This study	28099_1_321	ERS2855079	Stool	Malawi
This study	28099_1_322	ERS2855176	Stool	Malawi
This study	28099_1_323	ERS2855280	Stool	Malawi
This study	28099_1_325	ERS2855080	Stool	Malawi
This study	28099_1_326	ERS2855177	Stool	Malawi
This study	28099_1_327	ERS2855281	Stool	Malawi
This study	28099_1_329	ERS2855081	Stool	Malawi
This study	28099_1_330	ERS2855178	Stool	Malawi
This study	28099_1_331	ERS2855282	Stool	Malawi
This study	28099_1_333	ERS2855082	Stool	Malawi
This study	28099_1_334	ERS2855179	Stool	Malawi
This study	28099_1_335	ERS2855283	Stool	Malawi
This study	28099_1_337	ERS2855083	Stool	Malawi
This study	28099_1_338	ERS2855180	Stool	Malawi
This study	28099_1_339	ERS2855284	Stool	Malawi
This study	28099_1_34	ERS2855103	Stool	Malawi
This study	28099_1_341	ERS2855084	Stool	Malawi
This study	28099_1_342	ERS2855181	Stool	Malawi
This study	28099_1_343	ERS2855285	Stool	Malawi
This study	28099_1_345	ERS2855085	Stool	Malawi

Table 6.3: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
This study	28099_1_346	ERS2855182	Stool	Malawi
This study	28099_1_347	ERS2855286	Stool	Malawi
This study	28099_1_349	ERS2855086	Stool	Malawi
This study	28099_1_35	ERS2855199	Stool	Malawi
This study	28099_1_350	ERS2855183	Stool	Malawi
This study	28099_1_351	ERS2855287	Stool	Malawi
This study	28099_1_353	ERS2855087	Stool	Malawi
This study	28099_1_354	ERS2855184	Stool	Malawi
This study	28099_1_355	ERS2855288	Stool	Malawi
This study	28099_1_357	ERS2855088	Stool	Malawi
This study	28099_1_358	ERS2855185	Stool	Malawi
This study	28099_1_359	ERS2855289	Stool	Malawi
This study	28099_1_361	ERS2855089	Stool	Malawi
This study	28099_1_362	ERS2855186	Stool	Malawi
This study	28099_1_363	ERS2855290	Stool	Malawi
This study	28099_1_365	ERS2855090	Stool	Malawi
This study	28099_1_366	ERS2855187	Stool	Malawi
This study	28099_1_367	ERS2855291	Stool	Malawi
This study	28099_1_370	ERS2855188	Stool	Malawi
This study	28099_1_371	ERS2855292	Stool	Malawi
This study	28099_1_373	ERS2855092	Stool	Malawi
This study	28099_1_374	ERS2855189	Stool	Malawi
This study	28099_1_375	ERS2855261	Stool	Malawi
This study	28099_1_377	ERS2855093	Stool	Malawi
This study	28099_1_378	ERS2855190	Stool	Malawi
This study	28099_1_379	ERS2855262	Stool	Malawi
This study	28099_1_38	ERS2855104	Stool	Malawi
This study	28099_1_381	ERS2855094	Stool	Malawi
This study	28099_1_382	ERS2855191	Stool	Malawi
This study	28099_1_383	ERS2855263	Stool	Malawi
This study	28099_1_39	ERS2855200	Stool	Malawi
This study	28099_1_41	ERS2855009	Stool	Malawi
This study	28099_1_42	ERS2855106	Stool	Malawi
This study	28099_1_43	ERS2855201	Stool	Malawi
This study	28099_1_46	ERS2855107	Stool	Malawi
This study	28099_1_47	ERS2855202	Stool	Malawi
This study	28099_1_49	ERS2855011	Stool	Malawi
This study	28099_1_50	ERS2855108	Stool	Malawi
This study	28099_1_51	ERS2855203	Stool	Malawi
This study	28099_1_53	ERS2855012	Stool	Malawi
This study	28099_1_54	ERS2855109	Stool	Malawi
This study	28099_1_55	ERS2855204	Stool	Malawi
This study	28099_1_57	ERS2855013	Stool	Malawi
This study	28099_1_58	ERS2855110	Stool	Malawi
This study	28099_1_59	ERS2855205	Stool	Malawi
This study	28099_1_61	ERS2855014	Stool	Malawi
This study	28099_1_62	ERS2855111	Stool	Malawi
This study	28099_1_63	ERS2855206	Stool	Malawi
This study	28099_1_65	ERS2855015	Stool	Malawi
This study	28099_1_66	ERS2855112	Stool	Malawi
This study	28099_1_69	ERS2855016	Stool	Malawi
This study	28099_1_7	ERS2855193	Stool	Malawi
This study	28099_1_70	ERS2855113	Stool	Malawi
This study	28099_1_71	ERS2855208	Stool	Malawi
This study	28099_1_73	ERS2855017	Stool	Malawi
This study	28099_1_74	ERS2855114	Stool	Malawi
This study	28099_1_75	ERS2855209	Stool	Malawi
This study	28099_1_77	ERS2855040	Stool	Malawi
This study	28099_1_78	ERS2855115	Stool	Malawi
This study	28099_1_79	ERS2855210	Stool	Malawi
This study	28099_1_81	ERS2855041	Stool	Malawi
This study	28099_1_82	ERS2855116	Stool	Malawi

Table 6.3: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
This study	28099_1_83	ERS2855211	Stool	Malawi
This study	28099_1_85	ERS2855042	Stool	Malawi
This study	28099_1_86	ERS2855117	Stool	Malawi
This study	28099_1_87	ERS2855212	Stool	Malawi
This study	28099_1_89	ERS2855043	Stool	Malawi
This study	28099_1_90	ERS2855118	Stool	Malawi
This study	28099_1_91	ERS2855213	Stool	Malawi
This study	28099_1_93	ERS2855044	Stool	Malawi
This study	28099_1_94	ERS2855119	Stool	Malawi
This study	28099_1_95	ERS2855214	Stool	Malawi
This study	28099_1_98	ERS2855045	Stool	Malawi
This study	28099_1_99	ERS2855120	Stool	Malawi
This study	28099_2_101	ERS2855313	Stool	Malawi
This study	28099_2_105	ERS2855314	Stool	Malawi
This study	28099_2_109	ERS2855315	Stool	Malawi
This study	28099_2_113	ERS2855316	Stool	Malawi
This study	28099_2_117	ERS2855317	Stool	Malawi
This study	28099_2_12	ERS2855266	Stool	Malawi
This study	28099_2_121	ERS2855318	Stool	Malawi
This study	28099_2_126	ERS2855319	Stool	Malawi
This study	28099_2_130	ERS2855320	Stool	Malawi
This study	28099_2_138	ERS2855322	Stool	Malawi
This study	28099_2_142	ERS2855323	Stool	Malawi
This study	28099_2_146	ERS2855324	Stool	Malawi
This study	28099_2_150	ERS2855325	Stool	Malawi
This study	28099_2_154	ERS2855326	Stool	Malawi
This study	28099_2_158	ERS2855327	Stool	Malawi
This study	28099_2_16	ERS2855267	Stool	Malawi
This study	28099_2_162	ERS2855328	Stool	Malawi
This study	28099_2_166	ERS2855329	Stool	Malawi
This study	28099_2_170	ERS2855330	Stool	Malawi
This study	28099_2_174	ERS2855331	Stool	Malawi
This study	28099_2_178	ERS2855332	Stool	Malawi
This study	28099_2_182	ERS2855333	Stool	Malawi
This study	28099_2_186	ERS2855334	Stool	Malawi
This study	28099_2_190	ERS2855335	Stool	Malawi
This study	28099_2_194	ERS2855336	Stool	Malawi
This study	28099_2_198	ERS2855337	Stool	Malawi
This study	28099_2_206	ERS2855339	Stool	Malawi
This study	28099_2_210	ERS2855340	Stool	Malawi
This study	28099_2_215	ERS2855341	Stool	Malawi
This study	28099_2_219	ERS2855342	Stool	Malawi
This study	28099_2_223	ERS2855343	Stool	Malawi
This study	28099_2_227	ERS2855344	Stool	Malawi
This study	28099_2_231	ERS2855345	Stool	Malawi
This study	28099_2_235	ERS2855346	Stool	Malawi
This study	28099_2_239	ERS2855347	Stool	Malawi
This study	28099_2_24	ERS2855293	Stool	Malawi
This study	28099_2_243	ERS2855349	Stool	Malawi
This study	28099_2_247	ERS2855350	Stool	Malawi
This study	28099_2_251	ERS2855351	Stool	Malawi
This study	28099_2_255	ERS2855352	Stool	Malawi
This study	28099_2_259	ERS2855353	Stool	Malawi
This study	28099_2_263	ERS2855354	Stool	Malawi
This study	28099_2_283	ERS2855359	Stool	Malawi
This study	28099_2_287	ERS2855360	Stool	Malawi
This study	28099_2_291	ERS2855361	Stool	Malawi
This study	28099_2_295	ERS2855362	Stool	Malawi
This study	28099_2_299	ERS2855363	Stool	Malawi
This study	28099_2_3	ERS2855264	Stool	Malawi
This study	28099_2_304	ERS2855364	Stool	Malawi
This study	28099_2_308	ERS2855365	Stool	Malawi



Table 6.3: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
This study	28099_2_316	ERS2855224	Stool	Malawi
This study	28099_2_32	ERS2855295	Stool	Malawi
This study	28099_2_320	ERS2855348	Stool	Malawi
This study	28099_2_36	ERS2855296	Stool	Malawi
This study	28099_2_40	ERS2855297	Stool	Malawi
This study	28099_2_44	ERS2855298	Stool	Malawi
This study	28099_2_48	ERS2855299	Stool	Malawi
This study	28099_2_56	ERS2855301	Stool	Malawi
This study	28099_2_60	ERS2855302	Stool	Malawi
This study	28099_2_64	ERS2855303	Stool	Malawi
This study	28099_2_76	ERS2855306	Stool	Malawi
This study	28099_2_8	ERS2855265	Stool	Malawi
This study	28099_2_80	ERS2855307	Stool	Malawi
This study	28099_2_84	ERS2855308	Stool	Malawi
This study	28099_2_88	ERS2855309	Stool	Malawi
This study	28099_2_92	ERS2855310	Stool	Malawi
This study	28099_2_96	ERS2855311	Stool	Malawi
This study	28099_2_97	ERS2855312	Stool	Malawi
Ingle 2018	100269_aEPEC	ERR134513	Stool	Gambia
Ingle 2018	100383_aEPEC	ERR137807	Stool	Gambia
Ingle 2018	100446	ERR178176	Stool	Gambia
Ingle 2018	100554_aEPEC	ERR134514	Stool	Gambia
Ingle 2018	100600_aEPEC	ERR134515	Stool	Gambia
Ingle 2018	102010_aEPEC	ERR137808	Stool	Gambia
Ingle 2018	102014_aEPEC	ERR137809	Stool	Gambia
Ingle 2018	102298_aEPEC	ERR134516	Stool	Gambia
Ingle 2018	102328_aEPEC	ERR134517	Stool	Gambia
Ingle 2018	102366_aEPEC	ERR137810	Stool	Gambia
Ingle 2018	102485_aEPEC	ERR134518	Stool	Gambia
Ingle 2018	103151	ERR178192	Stool	Gambia
Ingle 2018	200135_aEPEC	ERR134519	Stool	Mali
Ingle 2018	200232	ERR178150	Stool	Mali
Ingle 2018	200439_aEPEC	ERR134520	Stool	Mali
Ingle 2018	200456_aEPEC	ERR137812	Stool	Mali
Ingle 2018	200499	ERR178148	Stool	Mali
Ingle 2018	200696	ERR178151	Stool	Mali
Ingle 2018	200708_aEPEC	ERR137782	Stool	Mali
Ingle 2018	200758_aEPEC	ERR137783	Stool	Mali
Ingle 2018	200781_aEPEC	ERR124658	Stool	Mali
Ingle 2018	200959_aEPEC	ERR137813	Stool	Mali
Ingle 2018	201191_aEPEC	ERR137814	Stool	Mali
Ingle 2018	201214_aEPEC	ERR134521	Stool	Mali
Ingle 2018	201350	ERR178216	Stool	Mali
Ingle 2018	201381_aEPEC	ERR137784	Stool	Mali
Ingle 2018	201488_aEPEC	ERR137815	Stool	Mali
Ingle 2018	201534_aEPEC	ERR134522	Stool	Mali
Ingle 2018	201589_aEPEC	ERR137816	Stool	Mali
Ingle 2018	202317_aEPEC	ERR137817	Stool	Mali
Ingle 2018	202374	ERR178152	Stool	Mali
Ingle 2018	202387	ERR178149	Stool	Mali
Ingle 2018	202423_aEPEC	ERR134523	Stool	Mali
Ingle 2018	202443_aEPEC	ERR134524	Stool	Mali
Ingle 2018	202453_aEPEC	ERR134525	Stool	Mali
Ingle 2018	202474	ERR178153	Stool	Mali
Ingle 2018	202521_aEPEC	ERR124659	Stool	Mali
Ingle 2018	202621_aEPEC	ERR137818	Stool	Mali
Ingle 2018	202833_aEPEC	ERR134526	Stool	Mali
Ingle 2018	202973_aEPEC	ERR134527	Stool	Mali
Ingle 2018	203470_aEPEC	ERR124660	Stool	Mali
Ingle 2018	204263_aEPEC	ERR124661	Stool	Mali
Ingle 2018	204302_aEPEC	ERR134528	Stool	Mali
Ingle 2018	300073	ERR178193	Stool	Mozambique

Table 6.3: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Ingle 2018	300086_aEPEC	ERR134529	Stool	Mozambique
Ingle 2018	300711_aEPEC	ERR134530	Stool	Mozambique
Ingle 2018	300795_aEPEC	ERR134531	Stool	Mozambique
Ingle 2018	300812_aEPEC	ERR137819	Stool	Mozambique
Ingle 2018	300814_aEPEC	ERR137820	Stool	Mozambique
Ingle 2018	302082	ERR178198	Stool	Mozambique
Ingle 2018	302302	ERR178154	Stool	Mozambique
Ingle 2018	302613	ERR178210	Stool	Mozambique
Ingle 2018	302619	ERR178211	Stool	Mozambique
Ingle 2018	302700	ERR178217	Stool	Mozambique
Ingle 2018	302701	ERR178212	Stool	Mozambique
Ingle 2018	302710	ERR178218	Stool	Mozambique
Ingle 2018	400549_aEPEC	ERR137785	Stool	Kenya
Ingle 2018	400654_aEPEC	ERR137786	Stool	Kenya
Ingle 2018	400714_aEPEC	ERR137787	Stool	Kenya
Ingle 2018	400896	ERR178177	Stool	Kenya
Ingle 2018	400897_aEPEC	ERR137821	Stool	Kenya
Ingle 2018	400998_aEPEC	ERR137789	Stool	Kenya
Ingle 2018	401082	ERR178178	Stool	Kenya
Ingle 2018	401117_aEPEC	ERR137822	Stool	Kenya
Ingle 2018	401174_aEPEC	ERR137823	Stool	Kenya
Ingle 2018	401250_aEPEC	ERR137790	Stool	Kenya
Ingle 2018	401352	ERR178199	Stool	Kenya
Ingle 2018	401363	ERR178179	Stool	Kenya
Ingle 2018	401480_aEPEC	ERR124657	Stool	Kenya
Ingle 2018	401553	ERR178155	Stool	Kenya
Ingle 2018	401596_aEPEC	ERR137791	Stool	Kenya
Ingle 2018	401686	ERR178200	Stool	Kenya
Ingle 2018	401709_aEPEC	ERR137824	Stool	Kenya
Ingle 2018	401886_aEPEC	ERR137792	Stool	Kenya
Ingle 2018	401907	ERR178201	Stool	Kenya
Ingle 2018	401938_aEPEC	ERR137793	Stool	Kenya
Ingle 2018	402048_aEPEC	ERR134532	Stool	Kenya
Ingle 2018	402058_aEPEC	ERR137825	Stool	Kenya
Ingle 2018	402074_aEPEC	ERR137794	Stool	Kenya
Ingle 2018	402078	ERR178180	Stool	Kenya
Ingle 2018	402097_aEPEC	ERR137826	Stool	Kenya
Ingle 2018	402099_aEPEC	ERR134533	Stool	Kenya
Ingle 2018	402138_aEPEC	ERR124662	Stool	Kenya
Ingle 2018	402227_aEPEC	ERR137827	Stool	Kenya
Ingle 2018	402248_aEPEC	ERR134534	Stool	Kenya
Ingle 2018	402403	ERR178181	Stool	Kenya
Ingle 2018	402480_aEPEC	ERR137795	Stool	Kenya
Ingle 2018	402605	ERR178194	Stool	Kenya
Ingle 2018	402617	ERR178156	Stool	Kenya
Ingle 2018	402635	ERR178157	Stool	Kenya
Ingle 2018	402654_aEPEC	ERR134535	Stool	Kenya
Ingle 2018	402696	ERR178202	Stool	Kenya
Ingle 2018	402743_aEPEC	ERR137796	Stool	Kenya
Ingle 2018	402767	ERR178203	Stool	Kenya
Ingle 2018	402770_aEPEC	ERR134536	Stool	Kenya
Ingle 2018	402780_aEPEC	ERR137797	Stool	Kenya
Ingle 2018	402794	ERR178204	Stool	Kenya
Ingle 2018	402798	ERR178182	Stool	Kenya
Ingle 2018	402837	ERR178183	Stool	Kenya
Ingle 2018	402842	ERR178205	Stool	Kenya
Ingle 2018	402898	ERR178184	Stool	Kenya
Ingle 2018	402924	ERR178158	Stool	Kenya
Ingle 2018	402977_aEPEC	ERR137798	Stool	Kenya
Ingle 2018	403066	ERR178159	Stool	Kenya
Ingle 2018	403096_aEPEC	ERR137799	Stool	Kenya
Ingle 2018	403128_aEPEC	ERR134537	Stool	Kenya

Table 6.3: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Ingle 2018	403308_aEPEC	ERR134538	Stool	Kenya
Ingle 2018	403523	ERR178206	Stool	Kenya
Ingle 2018	403726_aEPEC	ERR137800	Stool	Kenya
Ingle 2018	403728	ERR178161	Stool	Kenya
Ingle 2018	500094	ERR178207	Stool	India
Ingle 2018	500095	ERR178208	Stool	India
Ingle 2018	500193	ERR178213	Stool	India
Ingle 2018	500197_aEPEC	ERR137828	Stool	India
Ingle 2018	500275_aEPEC	ERR134539	Stool	India
Ingle 2018	500618_aEPEC	ERR134540	Stool	India
Ingle 2018	500858_aEPEC	ERR134541	Stool	India
Ingle 2018	500864_aEPEC	ERR134542	Stool	India
Ingle 2018	500989	ERR178164	Stool	India
Ingle 2018	501016	ERR178195	Stool	India
Ingle 2018	501029_aEPEC	ERR134543	Stool	India
Ingle 2018	503023	ERR178196	Stool	India
Ingle 2018	503028_aEPEC	ERR134544	Stool	India
Ingle 2018	503130_aEPEC	ERR137829	Stool	India
Ingle 2018	503225_aEPEC	ERR134545	Stool	India
Ingle 2018	503238_aEPEC	ERR137801	Stool	India
Ingle 2018	503256	ERR178197	Stool	India
Ingle 2018	503311_aEPEC	ERR124653	Stool	India
Ingle 2018	503320	ERR178219	Stool	India
Ingle 2018	503331	ERR178165	Stool	India
Ingle 2018	503459_aEPEC	ERR134546	Stool	India
Ingle 2018	503537_aEPEC	ERR124663	Stool	India
Ingle 2018	503662_aEPEC	ERR134547	Stool	India
Ingle 2018	503891_aEPEC	ERR137802	Stool	India
Ingle 2018	503947_aEPEC	ERR124654	Stool	India
Ingle 2018	504005_aEPEC	ERR137803	Stool	India
Ingle 2018	504180	ERR178166	Stool	India
Ingle 2018	504225_aEPEC	ERR134548	Stool	India
Ingle 2018	504300_aEPEC	ERR134549	Stool	India
Ingle 2018	504324	ERR178167	Stool	India
Ingle 2018	504449_aEPEC	ERR134550	Stool	India
Ingle 2018	504528	ERR178168	Stool	India
Ingle 2018	504647_aEPEC	ERR134551	Stool	India
Ingle 2018	504718	ERR178169	Stool	India
Ingle 2018	504821_aEPEC	ERR134552	Stool	India
Ingle 2018	504888_aEPEC	ERR134553	Stool	India
Ingle 2018	504925_aEPEC	ERR124664	Stool	India
Ingle 2018	505148	ERR178170	Stool	India
Ingle 2018	505393_aEPEC	ERR124655	Stool	India
Ingle 2018	505513_aEPEC	ERR124656	Stool	India
Ingle 2018	505545	ERR178171	Stool	India
Ingle 2018	602206	ERR178172	Stool	Bangladesh
Ingle 2018	602370_aEPEC	ERR134554	Stool	Bangladesh
Ingle 2018	700149	ERR178214	Stool	Pakistan
Ingle 2018	700337_aEPEC	ERR134555	Stool	Pakistan
Ingle 2018	700495_aEPEC	ERR134556	Stool	Pakistan
Ingle 2018	700851	ERR178173	Stool	Pakistan
Ingle 2018	700863	ERR178215	Stool	Pakistan
Ingle 2018	702161_aEPEC	ERR134558	Stool	Pakistan
Ingle 2018	702328	ERR178174	Stool	Pakistan
Ingle 2018	702566	ERR178175	Stool	Pakistan
Ingle 2018	702745_aEPEC	ERR137804	Stool	Pakistan
Ingle 2018	702797	ERR178186	Stool	Pakistan
Ingle 2018	702890_aEPEC	ERR137805	Stool	Pakistan
Ingle 2018	702898_aEPEC	ERR137806	Stool	Pakistan
Ingle 2018	702971	ERR178185	Stool	Pakistan
Ingle 2018	703063	ERR178209	Stool	Pakistan
Ingle 2018	703108	ERR178187	Stool	Pakistan

Table 6.3: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Ingle 2018	703128	ERR178188	Stool	Pakistan
Ingle 2018	703258_aEPEC	ERR134559	Stool	Pakistan
Ingle 2018	703273	ERR178191	Stool	Pakistan
Ingle 2018	703753	ERR178189	Stool	Pakistan
Ingle 2018	703975_aEPEC	ERR134560	Stool	Pakistan
Ingle 2018	G100788-1A	ERR175731	Stool	Gambia
Ingle 2018	G302544	ERR178226	Stool	Mozambique
Ingle 2018	G302551	ERR178225	Stool	Mozambique
Ingle 2018	G303212	ERR175730	Stool	Mozambique
Ingle 2018	G400792	ERR175725	Stool	Kenya
Ingle 2018	G400871	ERR175724	Stool	Kenya
Ingle 2018	G401436	ERR175727	Stool	Kenya
Ingle 2018	G401529	ERR178227	Stool	Kenya
Ingle 2018	G500007	ERR178223	Stool	India
Ingle 2018	G500297-1	ERR175733	Stool	India
Ingle 2018	G500407	ERR178221	Stool	India
Ingle 2018	G500830	ERR175728	Stool	India
Ingle 2018	G503854	ERR178224	Stool	India
Ingle 2018	G504540	ERR178222	Stool	India
Ingle 2018	G603423	ERR178228	Stool	Bangladesh
Ingle 2018	G702074-1	ERR175734	Stool	Pakistan
Ingle 2018	G702074-2	ERR175735	Stool	Pakistan
Ingle 2018	R203092-3A	ERR175736	Stool	Mali
Ingle 2018	R203092-3B	ERR175737	Stool	Mali
Ingle 2018	R302583-2A	ERR175738	Stool	Mozambique
Ingle 2018	R302583-2B	ERR175739	Stool	Mozambique
Ingle 2018	R402077	ERR175726	Stool	Kenya
Ingle 2018	R503696	ERR175729	Stool	India
Mentzer 2014	E_1003	ERR054711	Stool	Egypt
Mentzer 2014	E_1009	ERR054712	Stool	Egypt
Mentzer 2014	E_1018	ERR084463	Stool	Egypt
Mentzer 2014	E_1034	ERR052911	Stool	Egypt
Mentzer 2014	E_1057CFn	ERR119471	Stool	Egypt
Mentzer 2014	E_106	ERR054666	Stool	unknown
Mentzer 2014	E_1072CFn	ERR119472	Stool	Egypt
Mentzer 2014	E_1074	ERR052912	Stool	Egypt
Mentzer 2014	E_1085	ERR052913	Stool	Egypt
Mentzer 2014	E_1091	ERR052914	Stool	Egypt
Mentzer 2014	E_110	ERR054678	Stool	Bangladesh
Mentzer 2014	E_1101CFn	ERR119473	Stool	Egypt
Mentzer 2014	E_1102CFn	ERR119474	Stool	Egypt
Mentzer 2014	E_1111	ERR052915	Stool	Egypt
Mentzer 2014	E_1167CFn	ERR119475	Stool	Egypt
Mentzer 2014	E_1169CFn	ERR119476	Stool	Egypt
Mentzer 2014	E_1189	ERR052916	Stool	Egypt
Mentzer 2014	E_1193CFn	ERR119477	Stool	Egypt
Mentzer 2014	E_1242CFn	ERR119478	Stool	Egypt
Mentzer 2014	E_1245	ERR052917	Stool	Egypt
Mentzer 2014	E_1248CFn	ERR119479	Stool	Egypt
Mentzer 2014	E_1258CFn	ERR119480	Stool	Egypt
Mentzer 2014	E_126	ERR054679	Stool	unknown
Mentzer 2014	E_1264CFn	ERR119481	Stool	Egypt
Mentzer 2014	E_1281CFn	ERR119482	Stool	Egypt
Mentzer 2014	E_1282CFn	ERR119483	Stool	Egypt
Mentzer 2014	E_1285CFn	ERR119484	Stool	Egypt
Mentzer 2014	E_1287	ERR052918	Stool	Egypt
Mentzer 2014	E_129	ERR054680	Stool	Zaire
Mentzer 2014	E_1298	ERR052919	Stool	Egypt
Mentzer 2014	E_1316	ERR161000	Stool	Nepal
Mentzer 2014	E_133	ERR054681	Stool	unknown
Mentzer 2014	E_1334	ERR084464	Stool	China
Mentzer 2014	E_135	ERR054682	Stool	unknown

Table 6.3: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Mentzer 2014	E_1352CFn	ERR119485	Stool	Egypt
Mentzer 2014	E_1355CFn	ERR119486	Stool	Egypt
Mentzer 2014	E_1356CFn	ERR119487	Stool	Egypt
Mentzer 2014	E_1360_sec	ERR178234	Stool	Tunisia
Mentzer 2014	E_1362CFn	ERR119489	Stool	Egypt
Mentzer 2014	E_1363	ERR084466	Stool	Egypt
Mentzer 2014	E_1365CFn	ERR119490	Stool	Egypt
Mentzer 2014	E_1373	ERR052920	Stool	Indonesia
Mentzer 2014	E_1392	ERR052921	Stool	Indonesia
Mentzer 2014	E_1398CFn	ERR119491	Stool	Indonesia
Mentzer 2014	E_1407CFn	ERR119492	Stool	Mexico
Mentzer 2014	E_141	ERR054683	Stool	Burma
Mentzer 2014	E_1429tiny	ERR217371	Stool	Venezuela
Mentzer 2014	E_143	ERR054684	Stool	Japan
Mentzer 2014	E_1432G	ERR164830	Stool	Venezuela
Mentzer 2014	E_1432w	ERR164829	Stool	Venezuela
Mentzer 2014	E_1433	ERR084468	Stool	Morocco
Mentzer 2014	E_1460	ERR084469	Stool	Indonesia
Mentzer 2014	E_151	ERR054685	Stool	Japan
Mentzer 2014	E_1524	ERR084470	Stool	Argentina
Mentzer 2014	E_1525CFn	ERR119496	Stool	Argentina
Mentzer 2014	E_1526CFn	ERR119497	Stool	Argentina
Mentzer 2014	E_1527	ERR084471	Stool	Argentina
Mentzer 2014	E_1532CFn	ERR119498	Stool	Argentina
Mentzer 2014	E_1533CFn	ERR119499	Stool	Argentina
Mentzer 2014	E_1534CFn	ERR119500	Stool	Argentina
Mentzer 2014	E_1535	ERR084472	Stool	Argentina
Mentzer 2014	E_1541	ERR052922	Stool	Argentina
Mentzer 2014	E_1542CFn	ERR119501	Stool	Argentina
Mentzer 2014	E_1543CFn	ERR119502	Stool	Argentina
Mentzer 2014	E_1544CFn	ERR119503	Stool	Argentina
Mentzer 2014	E_1548	ERR052923	Stool	Argentina
Mentzer 2014	E_1556CFn	ERR119504	Stool	Argentina
Mentzer 2014	E_1561CFn	ERR119505	Stool	Argentina
Mentzer 2014	E_1564CFn	ERR119506	Stool	Argentina
Mentzer 2014	E_1573CFn	ERR119508	Stool	Argentina
Mentzer 2014	E_1574CFn	ERR119509	Stool	Argentina
Mentzer 2014	E_1576CFn	ERR119510	Stool	Argentina
Mentzer 2014	E_157CFn	ERR119507	Stool	Japan
Mentzer 2014	E_1580CFn	ERR119511	Stool	Argentina
Mentzer 2014	E_1581CFn	ERR119512	Stool	Argentina
Mentzer 2014	E_1582CFn	ERR119513	Stool	Argentina
Mentzer 2014	E_1585CFn	ERR119514	Stool	Argentina
Mentzer 2014	E_1586CFn	ERR119515	Stool	Argentina
Mentzer 2014	E_1587CFn	ERR119516	Stool	Argentina
Mentzer 2014	E_1592CFn	ERR119518	Stool	Argentina
Mentzer 2014	E_1593	ERR084473	Stool	Argentina
Mentzer 2014	E_1594CFn	ERR119519	Stool	Argentina
Mentzer 2014	E_1596CFn	ERR119520	Stool	Argentina
Mentzer 2014	E_1597CFn	ERR119521	Stool	Argentina
Mentzer 2014	E_1599CFn	ERR119522	Stool	Argentina
Mentzer 2014	E_159CFn	ERR119517	Stool	Japan
Mentzer 2014	E_1600CFn	ERR119524	Stool	Argentina
Mentzer 2014	E_1604CFn	ERR119525	Stool	Argentina
Mentzer 2014	E_1607CFn	ERR119526	Stool	Argentina
Mentzer 2014	E_1609CFn	ERR119527	Stool	Argentina
Mentzer 2014	E_1611CFn	ERR119528	Stool	Argentina
Mentzer 2014	E_1615CFn	ERR119529	Stool	Argentina
Mentzer 2014	E_1616CFn	ERR119530	Stool	Argentina
Mentzer 2014	E_1617CFn	ERR119531	Stool	Argentina
Mentzer 2014	E_1620	ERR084474	Stool	Argentina
Mentzer 2014	E_1623CFn	ERR119532	Stool	Indonesia

Table 6.3: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Mentzer 2014	E_1624	ERR052924	Stool	Indonesia
Mentzer 2014	E_1625CFn	ERR119533	Stool	Indonesia
Mentzer 2014	E_1628CFn	ERR119534	Stool	Indonesia
Mentzer 2014	E_1634CFn	ERR119535	Stool	Indonesia
Mentzer 2014	E_1635	ERR084475	Stool	Indonesia
Mentzer 2014	E_1637CFn	ERR119536	Stool	Indonesia
Mentzer 2014	E_1638CFn	ERR119537	Stool	Indonesia
Mentzer 2014	E_1640CFn	ERR119538	Stool	Indonesia
Mentzer 2014	E_1641CFn	ERR119539	Stool	Indonesia
Mentzer 2014	E_1642CFn	ERR119540	Stool	Indonesia
Mentzer 2014	E_1646	ERR052925	Stool	Indonesia
Mentzer 2014	E_1647CFn	ERR119541	Stool	Indonesia
Mentzer 2014	E_1648CFn	ERR119542	Stool	Indonesia
Mentzer 2014	E_1649	ERR084476	Stool	Indonesia
Mentzer 2014	E_1650CFn	ERR119543	Stool	Indonesia
Mentzer 2014	E_1654	ERR054665	Stool	Indonesia
Mentzer 2014	E_1657	ERR084477	Stool	Indonesia
Mentzer 2014	E_1659CFn	ERR119544	Stool	Indonesia
Mentzer 2014	E_1661CFn	ERR119545	Stool	Indonesia
Mentzer 2014	E_1666CFn	ERR119546	Stool	Indonesia
Mentzer 2014	E_1667	ERR084478	Stool	Indonesia
Mentzer 2014	E_1673CFn	ERR119548	Stool	Indonesia
Mentzer 2014	E_1674CFn	ERR119549	Stool	Indonesia
Mentzer 2014	E_1679sec	ERR217373	Stool	Indonesia
Mentzer 2014	E_167CFn	ERR119547	Stool	Japan
Mentzer 2014	E_1682CFn	ERR119551	Stool	Indonesia
Mentzer 2014	E_1684CFn	ERR119552	Stool	Indonesia
Mentzer 2014	E_1690CFn	ERR119553	Stool	Indonesia
Mentzer 2014	E_1712CFn	ERR119554	Stool	Bangladesh
Mentzer 2014	E_1716	ERR052926	Stool	Bangladesh
Mentzer 2014	E_1724	ERR084479	Stool	Bangladesh
Mentzer 2014	E_1735CFn	ERR119555	Stool	Bangladesh
Mentzer 2014	E_1736CFn	ERR119556	Stool	Bangladesh
Mentzer 2014	E_1739CFn	ERR119557	Stool	Bangladesh
Mentzer 2014	E_1741	ERR084480	Stool	Bangladesh
Mentzer 2014	E_1744CFn	ERR119558	Stool	Bangladesh
Mentzer 2014	E_1750	ERR084481	Stool	Bangladesh
Mentzer 2014	E_1752CFn	ERR119559	Stool	Bangladesh
Mentzer 2014	E_1760	ERR084482	Stool	Bangladesh
Mentzer 2014	E_1779	ERR052927	Stool	Bangladesh
Mentzer 2014	E_1784	ERR052928	Stool	Bangladesh
Mentzer 2014	E_1795	ERR084483	Stool	Bangladesh
Mentzer 2014	E_1797	ERR084484	Stool	Bangladesh
Mentzer 2014	E_1841	ERR084485	Stool	Bangladesh
Mentzer 2014	E_1871CFn	ERR119560	Stool	Bangladesh
Mentzer 2014	E_1883	ERR084486	Stool	Bangladesh
Mentzer 2014	E_1918	ERR084487	Stool	Bangladesh
Mentzer 2014	E_1939	ERR084488	Stool	Bangladesh
Mentzer 2014	E_1947	ERR084489	Stool	Bangladesh
Mentzer 2014	E_1961CFn	ERR119561	Stool	Bangladesh
Mentzer 2014	E_1994	ERR084490	Stool	Bangladesh
Mentzer 2014	E_2088	ERR084491	Stool	Bangladesh
Mentzer 2014	E_2092	ERR089723	Stool	Bangladesh
Mentzer 2014	E_21	ERR054667	Stool	Bangladesh
Mentzer 2014	E_2108CFn	ERR119562	Stool	Bangladesh
Mentzer 2014	E_2110CFn	ERR119563	Stool	Bangladesh
Mentzer 2014	E_2118	ERR089724	Stool	Bangladesh
Mentzer 2014	E_2121CFn	ERR119564	Stool	Bangladesh
Mentzer 2014	E_2131	ERR089725	Stool	Bangladesh
Mentzer 2014	E_2185	ERR089726	Stool	Bolivia
Mentzer 2014	E_220	ERR054686	Stool	Japan
Mentzer 2014	E_2219	ERR089727	Stool	Bolivia

Table 6.3: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Mentzer 2014	E_222CFn	ERR119380	Stool	Japan
Mentzer 2014	E_223CFn	ERR119381	Stool	Japan
Mentzer 2014	E_224	ERR054687	Stool	Japan
Mentzer 2014	E_2256	ERR089728	Stool	Thailand
Mentzer 2014	E_2339	ERR089729	Stool	Bolivia
Mentzer 2014	E_2347	ERR089730	Stool	Bolivia
Mentzer 2014	E_2348	ERR089731	Stool	Bolivia
Mentzer 2014	E_2362_sec	ERR178236	Stool	Bolivia
Mentzer 2014	E_237	ERR054688	Stool	Japan
Mentzer 2014	E_2370sec	ERR217375	Stool	Japan
Mentzer 2014	E_2386	ERR089732	Stool	Bolivia
Mentzer 2014	E_2397	ERR089733	Stool	Bolivia
Mentzer 2014	E_239CFn	ERR119382	Stool	Japan
Mentzer 2014	E_2439	ERR164832	Stool	Bolivia
Mentzer 2014	E_251	ERR054689	Stool	Japan
Mentzer 2014	E_263CFn	ERR119383	Stool	Japan
Mentzer 2014	E_272	ERR054690	Stool	Japan
Mentzer 2014	E_28	ERR054668	Stool	Bangladesh
Mentzer 2014	E_2980	ERR089734	Stool	Bangladesh
Mentzer 2014	E_2981_sec	ERR178239	Stool	Bangladesh
Mentzer 2014	E_329CFn	ERR119384	Stool	Mexico
Mentzer 2014	E_330CFn	ERR119385	Stool	Mexico
Mentzer 2014	E_333	ERR049162	Stool	Mexico
Mentzer 2014	E_335CFn	ERR119386	Stool	Mexico
Mentzer 2014	E_336CFn	ERR119387	Stool	Mexico
Mentzer 2014	E_340CFn	ERR119388	Stool	Mexico
Mentzer 2014	E_343CFn	ERR119389	Stool	Mexico
Mentzer 2014	E_344	ERR049163	Stool	Mexico
Mentzer 2014	E_351CFn	ERR119390	Stool	Mexico
Mentzer 2014	E_354CFn	ERR119391	Stool	Mexico
Mentzer 2014	E_356CFn	ERR119392	Stool	Mexico
Mentzer 2014	E_36	ERR054669	Stool	Bangladesh
Mentzer 2014	E_360CFn	ERR119393	Stool	Mexico
Mentzer 2014	E_361CFn	ERR119394	Stool	Mexico
Mentzer 2014	E_370	ERR049164	Stool	Guatemala
Mentzer 2014	E_390CFn	ERR119395	Stool	Guatemala
Mentzer 2014	E_391CFn	ERR119396	Stool	Guatemala
Mentzer 2014	E_399CFn	ERR119397	Stool	Guatemala
Mentzer 2014	E_405CFn	ERR119398	Stool	Guatemala
Mentzer 2014	E_415CFn	ERR119399	Stool	Guatemala
Mentzer 2014	E_416	ERR049165	Stool	Guatemala
Mentzer 2014	E_425CFn	ERR119400	Stool	Guatemala
Mentzer 2014	E_445CFn	ERR119401	Stool	Guatemala
Mentzer 2014	E_45	ERR054670	Stool	Bangladesh
Mentzer 2014	E_451CFn	ERR119402	Stool	Guatemala
Mentzer 2014	E_471	ERR049166	Stool	Guatemala
Mentzer 2014	E_5089	ERR164833	Stool	Bangladesh
Mentzer 2014	E_509	ERR178229	Stool	Mexico
Mentzer 2014	E_513CFn	ERR119404	Stool	Mexico
Mentzer 2014	E_517	ERR049167	Stool	Mexico
Mentzer 2014	E_519CFn	ERR119405	Stool	Mexico
Mentzer 2014	E_520CFn	ERR119406	Stool	Mexico
Mentzer 2014	E_523CFn	ERR119407	Stool	Mexico
Mentzer 2014	E_527CFn	ERR119408	Stool	Mexico
Mentzer 2014	E_528CFn	ERR119409	Stool	Mexico
Mentzer 2014	E_529CFn	ERR119410	Stool	Mexico
Mentzer 2014	E_54	ERR049158	Stool	Bangladesh
Mentzer 2014	E_554	ERR049168	Stool	Mexico
Mentzer 2014	E_562	ERR049169	Stool	Mexico
Mentzer 2014	E_563	ERR049170	Stool	Mexico
Mentzer 2014	E_604CFn	ERR119411	Stool	Guatemala
Mentzer 2014	E_616CFn	ERR119412	Stool	Guatemala

Table 6.3: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Mentzer 2014	E_618CFn	ERR119413	Stool	Guatemala
Mentzer 2014	E_620CFn	ERR119414	Stool	Guatemala
Mentzer 2014	E_621	ERR178230	Stool	Guatemala
Mentzer 2014	E_622CFn	ERR119416	Stool	Guatemala
Mentzer 2014	E_626CFn	ERR119417	Stool	Guatemala
Mentzer 2014	E_628CFn	ERR119418	Stool	Guatemala
Mentzer 2014	E_632	ERR049171	Stool	Guatemala
Mentzer 2014	E_636	ERR049172	Stool	Guatemala
Mentzer 2014	E_645CFn	ERR119419	Stool	Guatemala
Mentzer 2014	E_655	ERR049173	Stool	Guatemala
Mentzer 2014	E_658CFn	ERR119420	Stool	Guatemala
Mentzer 2014	E_659CFn	ERR119421	Stool	Guatemala
Mentzer 2014	E_66	ERR054671	Stool	Bangladesh
Mentzer 2014	E_662CFn	ERR119422	Stool	Guatemala
Mentzer 2014	E_70	ERR049159	Stool	Bangladesh
Mentzer 2014	E_703CFn	ERR119423	Stool	Guatemala
Mentzer 2014	E_704CFn	ERR119424	Stool	Guatemala
Mentzer 2014	E_705CFn	ERR119425	Stool	Guatemala
Mentzer 2014	E_71	ERR049160	Stool	Bangladesh
Mentzer 2014	E_710	ERR178231	Stool	Guatemala
Mentzer 2014	E_74	ERR054672	Stool	Bangladesh
Mentzer 2014	E_79	ERR054673	Stool	Bangladesh
Mentzer 2014	E_8	ERR049156	Stool	Bangladesh
Mentzer 2014	E_806	ERR054691	Stool	Guatemala
Mentzer 2014	E_810	ERR054692	Stool	Guatemala
Mentzer 2014	E_811	ERR178232	Stool	Guatemala
Mentzer 2014	E_812	ERR054693	Stool	Guatemala
Mentzer 2014	E_816	ERR054694	Stool	Guatemala
Mentzer 2014	E_818CFn	ERR119428	Stool	Guatemala
Mentzer 2014	E_819CFn	ERR119429	Stool	Guatemala
Mentzer 2014	E_821CFn	ERR119430	Stool	Guatemala
Mentzer 2014	E_822	ERR054695	Stool	Guatemala
Mentzer 2014	E_828CFn	ERR119431	Stool	Guatemala
Mentzer 2014	E_833CFn	ERR119432	Stool	Guatemala
Mentzer 2014	E_841CFn	ERR119433	Stool	Guatemala
Mentzer 2014	E_842	ERR054696	Stool	Guatemala
Mentzer 2014	E_85	ERR054674	Stool	Bangladesh
Mentzer 2014	E_855CFn	ERR119434	Stool	Guatemala
Mentzer 2014	E_856	ERR054697	Stool	Guatemala
Mentzer 2014	E_858	ERR054698	Stool	Guatemala
Mentzer 2014	E_860CFn	ERR119435	Stool	Guatemala
Mentzer 2014	E_863	ERR049174	Stool	Guatemala
Mentzer 2014	E_865CFn	ERR119436	Stool	Guatemala
Mentzer 2014	E_867	ERR054699	Stool	Guatemala
Mentzer 2014	E_87	ERR054675	Stool	Bangladesh
Mentzer 2014	E_871	ERR054700	Stool	Guatemala
Mentzer 2014	E_873CFn	ERR119437	Stool	Guatemala
Mentzer 2014	E_876CFn	ERR119438	Stool	Guatemala
Mentzer 2014	E_877	ERR049175	Stool	Guatemala
Mentzer 2014	E_879	ERR049176	Stool	Guatemala
Mentzer 2014	E_88	ERR049161	Stool	Bangladesh
Mentzer 2014	E_881CFn	ERR119439	Stool	Guatemala
Mentzer 2014	E_882	ERR054701	Stool	Guatemala
Mentzer 2014	E_883	ERR054702	Stool	Guatemala
Mentzer 2014	E_884CFn	ERR119441	Stool	Guatemala
Mentzer 2014	E_885CFn	ERR119442	Stool	Guatemala
Mentzer 2014	E_887CFn	ERR119443	Stool	Guatemala
Mentzer 2014	E_888CFn	ERR119444	Stool	Guatemala
Mentzer 2014	E_890CFn	ERR119445	Stool	Guatemala
Mentzer 2014	E_891CFn	ERR119446	Stool	Guatemala
Mentzer 2014	E_892CFn	ERR119447	Stool	Guatemala
Mentzer 2014	E_893CFn	ERR119448	Stool	Guatemala



Table 6.3: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Mentzer 2014	E_895	ERR054703	Stool	Guatemala
Mentzer 2014	E_897	ERR049177	Stool	Guatemala
Mentzer 2014	E_898CFn	ERR119449	Stool	Guatemala
Mentzer 2014	E_899CFn	ERR119450	Stool	Guatemala
Mentzer 2014	E_900	ERR054704	Stool	Guatemala
Mentzer 2014	E_901	ERR054705	Stool	Guatemala
Mentzer 2014	E_903	ERR054706	Stool	Guatemala
Mentzer 2014	E_907	ERR054707	Stool	Guatemala
Mentzer 2014	E_908CFn	ERR119452	Stool	Guatemala
Mentzer 2014	E_916	ERR049178	Stool	Guatemala
Mentzer 2014	E_917	ERR049179	Stool	Guatemala
Mentzer 2014	E_920	ERR054708	Stool	Guatemala
Mentzer 2014	E_924CFn	ERR119453	Stool	Guatemala
Mentzer 2014	E_925	ERR052905	Stool	Guatemala
Mentzer 2014	E_927	ERR052906	Stool	Egypt
Mentzer 2014	E_928	ERR054709	Stool	Egypt
Mentzer 2014	E_934CFn	ERR119455	Stool	Egypt
Mentzer 2014	E_935CFn	ERR119456	Stool	Egypt
Mentzer 2014	E_936CFn	ERR119457	Stool	Egypt
Mentzer 2014	E_938	ERR052907	Stool	Egypt
Mentzer 2014	E_939CFn	ERR119458	Stool	Egypt
Mentzer 2014	E_940	ERR052908	Stool	Egypt
Mentzer 2014	E_941CFn	ERR119459	Stool	Egypt
Mentzer 2014	E_943	ERR054710	Stool	Egypt
Mentzer 2014	E_944CFn	ERR119460	Stool	Egypt
Mentzer 2014	E_945CFn	ERR119461	Stool	Egypt
Mentzer 2014	E_947CFn	ERR119462	Stool	Egypt
Mentzer 2014	E_949CFn	ERR119463	Stool	Egypt
Mentzer 2014	E_952CFn	ERR119464	Stool	Egypt
Mentzer 2014	E_953	ERR052909	Stool	Egypt
Mentzer 2014	E_955CFn	ERR119465	Stool	Egypt
Mentzer 2014	E_956CFn	ERR119466	Stool	Egypt
Mentzer 2014	E_957	ERR164828	Stool	Egypt
Mentzer 2014	E_97	ERR054676	Stool	Bangladesh
Mentzer 2014	E_978CFn	ERR119468	Stool	Egypt
Mentzer 2014	E_986	ERR052910	Stool	Egypt
Mentzer 2014	E_99	ERR054677	Stool	Bangladesh
Mentzer 2014	E_995	ERR160999	Stool	Egypt
Mentzer 2014	E_996CFn	ERR119469	Stool	Egypt
Mentzer 2014	E_998CFn	ERR119470	Stool	Egypt
Mentzer 2014	E160CFn	ERR119523	Stool	Japan
Mentzer 2014	E2367CFn	ERR119566	Stool	Bolivia
Mentzer 2014	E2371CFn	ERR119568	Stool	Bolivia
Mentzer 2014	E2377CFn	ERR119569	Stool	Bolivia
Mentzer 2014	E2388CFn	ERR119570	Stool	Bolivia
Mentzer 2014	E2392CFn	ERR119571	Stool	Bolivia
Mentzer 2014	E2393CFn	ERR119572	Stool	Bolivia
Mentzer 2014	E2395CFn	ERR119573	Stool	Bolivia
Mentzer 2014	E2404CFn	ERR119574	Stool	Bolivia
Mentzer 2014	E2405CFn	ERR119575	Stool	Bolivia
Mentzer 2014	E3015CFn	ERR119577	Stool	Egypt
Mentzer 2014	E4134CFn	ERR119578	Stool	Israel
Mentzer 2014	E5049	ERR089738	Stool	India
Mentzer 2014	E5051	ERR089739	Stool	India
Mentzer 2014	E5052	ERR089740	Stool	India
Mentzer 2014	E5080	ERR089741	Stool	Bangladesh
Mentzer 2014	E5081	ERR089742	Stool	Bangladesh
Mentzer 2014	E5082	ERR089743	Stool	Bangladesh
Mentzer 2014	E5084	ERR089744	Stool	Bangladesh
Mentzer 2014	E5085	ERR089745	Stool	Bangladesh
Mentzer 2014	E5086	ERR089746	Stool	Bangladesh
Mentzer 2014	E5087	ERR089747	Stool	Bangladesh

Table 6.3: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Mentzer 2014	E5088	ERR089748	Stool	Bangladesh
Mentzer 2014	ILBEcoli5442571	ERR279354	Stool	Kenya
Mentzer 2014	ILBEcoli5442572	ERR279355	Stool	Kenya
Mentzer 2014	ILBEcoli5442573	ERR279356	Stool	Kenya
Mentzer 2014	ILBEcoli5442574	ERR279357	Stool	Kenya
Mentzer 2014	ILBEcoli5442575	ERR279358	Stool	Kenya
Mentzer 2014	ILBEcoli5442576	ERR279359	Stool	Kenya
Mentzer 2014	ILBEcoli5442577	ERR279360	Stool	Kenya
Mentzer 2014	ILBEcoli5442578	ERR279361	Stool	Kenya
Mentzer 2014	ILBEcoli5442579	ERR279362	Stool	Kenya
Mentzer 2014	ILBEcoli5442580	ERR279363	Stool	Kenya
Mentzer 2014	ILBEcoli5442581	ERR279364	Stool	Kenya
Mentzer 2014	ILBEcoli5442582	ERR279365	Stool	Kenya
Mentzer 2014	ILBEcoli5442583	ERR279366	Stool	Kenya
Mentzer 2014	ILBEcoli5442587	ERR279370	Stool	Guinea Bissau
Mentzer 2014	ILBEcoli5442588	ERR279371	Stool	Guinea Bissau
Mentzer 2014	ILBEcoli5442589	ERR279372	Stool	Guinea Bissau
Mentzer 2014	ILBEcoli5442590	ERR279373	Stool	Guinea Bissau
Musicha 2017	3487STDY6036382	ERR926351	Blood	Malawi
Musicha 2017	3487STDY6036383	ERR926352	Blood	Malawi
Musicha 2017	3487STDY6036384	ERR926353	Blood	Malawi
Musicha 2017	3487STDY6036385	ERR926354	CSF	Malawi
Musicha 2017	3487STDY6036386	ERR926355	Blood	Malawi
Musicha 2017	3487STDY6036387	ERR926356	Blood	Malawi
Musicha 2017	3487STDY6036388	ERR926357	Blood	Malawi
Musicha 2017	3487STDY6036389	ERR926358	Blood	Malawi
Musicha 2017	3487STDY6036390	ERR926359	CSF	Malawi
Musicha 2017	3487STDY6036391	ERR926360	Blood	Malawi
Musicha 2017	3487STDY6036392	ERR926361	Blood	Malawi
Musicha 2017	3487STDY6036393	ERR926362	Blood	Malawi
Musicha 2017	3487STDY6036394	ERR926363	Blood	Malawi
Musicha 2017	3487STDY6036395	ERR926364	Blood	Malawi
Musicha 2017	3487STDY6036396	ERR926365	CSF	Malawi
Musicha 2017	3487STDY6036397	ERR926366	Blood	Malawi
Musicha 2017	3487STDY6036398	ERR926367	Blood	Malawi
Musicha 2017	3487STDY6036399	ERR926368	Blood	Malawi
Musicha 2017	3487STDY6036400	ERR926369	CSF	Malawi
Musicha 2017	3487STDY6036403	ERR926372	CSF	Malawi
Musicha 2017	3487STDY6036404	ERR926373	Blood	Malawi
Musicha 2017	3487STDY6036405	ERR926374	CSF	Malawi
Musicha 2017	3487STDY6036406	ERR926375	Blood	Malawi
Musicha 2017	3487STDY6036407	ERR926376	CSF	Malawi
Musicha 2017	3487STDY6036408	ERR926377	Blood	Malawi
Musicha 2017	3487STDY6036409	ERR926378	Blood	Malawi
Musicha 2017	3487STDY6036410	ERR926379	Blood	Malawi
Musicha 2017	3487STDY6036411	ERR926380	Blood	Malawi
Musicha 2017	3487STDY6036412	ERR926381	Blood	Malawi
Musicha 2017	3487STDY6036413	ERR926382	CSF	Malawi
Musicha 2017	3487STDY6036414	ERR926383	Blood	Malawi
Musicha 2017	3487STDY6036415	ERR926384	CSF	Malawi
Musicha 2017	3487STDY6036416	ERR926385	CSF	Malawi
Musicha 2017	3487STDY6036417	ERR926386	CSF	Malawi
Musicha 2017	3487STDY6036418	ERR926387	CSF	Malawi
Musicha 2017	3487STDY6036419	ERR926388	CSF	Malawi
Musicha 2017	3487STDY6036420	ERR926389	Blood	Malawi
Musicha 2017	3487STDY6036421	ERR926390	CSF	Malawi
Musicha 2017	3487STDY6036422	ERR926391	CSF	Malawi
Musicha 2017	3487STDY6036423	ERR926392	CSF	Malawi
Musicha 2017	3487STDY6036424	ERR926393	Blood	Malawi
Musicha 2017	3487STDY6036425	ERR926394	Blood	Malawi
Musicha 2017	3487STDY6036426	ERR926395	Blood	Malawi
Musicha 2017	3487STDY6036427	ERR926396	Blood	Malawi
Musicha 2017	3487STDY6036428	ERR926397	Blood	Malawi

Table 6.3: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Musicha 2017	3487STDY6036429	ERR926397	Blood	Malawi
Musicha 2017	3487STDY6036430	ERR926398	Blood	Malawi
Musicha 2017	3487STDY6036431	ERR926399	Blood	Malawi
Musicha 2017	3487STDY6036432	ERR926400	Blood	Malawi
Musicha 2017	3487STDY6036433	ERR926401	CSF	Malawi
Musicha 2017	3487STDY6036434	ERR926402	CSF	Malawi
Musicha 2017	3487STDY6036435	ERR926403	CSF	Malawi
Musicha 2017	3487STDY6036436	ERR926404	Blood	Malawi
Musicha 2017	3487STDY6036437	ERR926405	Blood	Malawi
Musicha 2017	3487STDY6036438	ERR926406	Blood	Malawi
Musicha 2017	3487STDY6036440	ERR926408	CSF	Malawi
Musicha 2017	3487STDY6036441	ERR926409	Blood	Malawi
Musicha 2017	3487STDY6036443	ERR926411	Blood	Malawi
Musicha 2017	3487STDY6036444	ERR926412	Blood	Malawi
Musicha 2017	3487STDY6036445	ERR926413	RS	Malawi
Musicha 2017	3487STDY6036446	ERR971989	Blood	Malawi
Musicha 2017	3487STDY6036447	ERR926414	CSF	Malawi
Musicha 2017	3487STDY6036448	ERR926415	Blood	Malawi
Musicha 2017	3487STDY6036449	ERR926416	Blood	Malawi
Musicha 2017	3487STDY6036450	ERR926417	RS	Malawi
Musicha 2017	3487STDY6036451	ERR926418	CSF	Malawi
Musicha 2017	3487STDY6036452	ERR926419	Blood	Malawi
Musicha 2017	3487STDY6036453	ERR926420	Blood	Malawi
Musicha 2017	3487STDY6036454	ERR971990	Blood	Malawi
Musicha 2017	3487STDY6036455	ERR926421	CSF	Malawi
Musicha 2017	3487STDY6036456	ERR971991	Blood	Malawi
Musicha 2017	3487STDY6036457	ERR926422	RS	Malawi
Musicha 2017	3487STDY6036458	ERR926423	Blood	Malawi
Musicha 2017	3487STDY6036460	ERR926425	CSF	Malawi
Musicha 2017	3487STDY6036461	ERR926426	Blood	Malawi
Musicha 2017	3487STDY6036462	ERR971992	Blood	Malawi
Musicha 2017	3487STDY6036463	ERR926427	CSF	Malawi
Musicha 2017	3487STDY6036464	ERR926428	Blood	Malawi
Musicha 2017	3487STDY6036465	ERR926429	Blood	Malawi
Musicha 2017	3487STDY6036466	ERR926430	RS	Malawi
Musicha 2017	3487STDY6036467	ERR926431	RS	Malawi
Musicha 2017	3487STDY6036468	ERR926432	Blood	Malawi
Musicha 2017	3487STDY6036470	ERR926434	Blood	Malawi
Musicha 2017	3487STDY6036471	ERR926435	RS	Malawi
Musicha 2017	3487STDY6036473	ERR971994	RS	Malawi
Musicha 2017	3487STDY6036486	ERR926444	Blood	Malawi
Musicha 2017	3487STDY6036506	ERR971962	Blood	Malawi
Musicha 2017	3487STDY6036508	ERR971963	Blood	Malawi
Musicha 2017	3487STDY6036519	ERR972008	Blood	Malawi
Musicha 2017	3487STDY6036520	ERR971966	Blood	Malawi
Musicha 2017	3487STDY6036526	ERR971968	Blood	Malawi
Musicha 2017	3487STDY6036533	ERR971970	Blood	Malawi
Musicha 2017	3487STDY6036547	ERR971979	Blood	Malawi
Musicha 2017	3487STDY6036565	ERR971987	RS	Malawi
Runchaeron 2017	3898STDY6199571	ERR1218581	Urine	Thailand
Runchaeron 2017	3898STDY6199572	ERR1218582	Urine	Thailand
Runchaeron 2017	3898STDY6199573	ERR1218583	Urine	Thailand
Runchaeron 2017	3898STDY6199574	ERR1218584	Sputum	Thailand
Runchaeron 2017	3898STDY6199575	ERR1218585	Urine	Thailand
Runchaeron 2017	3898STDY6199576	ERR1218534	Blood	Thailand
Runchaeron 2017	3898STDY6199577	ERR1218586	Pus	Thailand
Runchaeron 2017	3898STDY6199578	ERR1218587	Blood	Thailand
Runchaeron 2017	3898STDY6199579	ERR1218588	Urine	Thailand
Runchaeron 2017	3898STDY6199580	ERR1218589	Urine	Thailand
Runchaeron 2017	3898STDY6199581	ERR1218590	Pus	Thailand
Runchaeron 2017	3898STDY6199582	ERR1218591	Pus	Thailand
Runchaeron 2017	3898STDY6199583	ERR1218592	Urine	Thailand

Table 6.3: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Runchaeron 2017	3898STDY6199584	ERR1218593	Urine	Thailand
Runchaeron 2017	3898STDY6199585	ERR1218594	Urine	Thailand
Runchaeron 2017	3898STDY6199586	ERR1218595	Blood	Thailand
Runchaeron 2017	3898STDY6199587	ERR1218596	Pus	Thailand
Runchaeron 2017	3898STDY6199588	ERR1218597	Urine	Thailand
Runchaeron 2017	3898STDY6199589	ERR1218535	Blood	Thailand
Runchaeron 2017	3898STDY6199590	ERR1218598	Urine	Thailand
Runchaeron 2017	3898STDY6199591	ERR1218599	Pus	Thailand
Runchaeron 2017	3898STDY6199592	ERR1218600	Urine	Thailand
Runchaeron 2017	3898STDY6199593	ERR1218536	Blood	Thailand
Runchaeron 2017	3898STDY6199594	ERR1218601	Urine	Thailand
Runchaeron 2017	3898STDY6199595	ERR1218602	Blood	Thailand
Runchaeron 2017	3898STDY6199596	ERR1218537	Pus	Thailand
Runchaeron 2017	3898STDY6199597	ERR1218538	Pus	Thailand
Runchaeron 2017	3898STDY6199598	ERR1218603	Pus	Thailand
Runchaeron 2017	3898STDY6199599	ERR1218539	Urine	Thailand
Runchaeron 2017	3898STDY6199600	ERR1218604	Pus	Thailand
Runchaeron 2017	3898STDY6199601	ERR1218540	Pus	Thailand
Runchaeron 2017	3898STDY6199602	ERR1218605	Urine	Thailand
Runchaeron 2017	3898STDY6199603	ERR1218606	Urine	Thailand
Runchaeron 2017	3898STDY6199604	ERR1218607	Urine	Thailand
Runchaeron 2017	3898STDY6199605	ERR1218608	Blood	Thailand
Runchaeron 2017	3898STDY6199606	ERR1218609	Pus	Thailand
Runchaeron 2017	3898STDY6199607	ERR1218610	Pus	Thailand
Runchaeron 2017	3898STDY6199608	ERR1218541	Urine	Thailand
Runchaeron 2017	3898STDY6199609	ERR1218611	Blood	Thailand
Runchaeron 2017	3898STDY6199610	ERR1218542	Pus	Thailand
Runchaeron 2017	3898STDY6199611	ERR1218612	Urine	Thailand
Runchaeron 2017	3898STDY6199612	ERR1218613	Pus	Thailand
Runchaeron 2017	3898STDY6199613	ERR1218614	Blood	Thailand
Runchaeron 2017	3898STDY6199614	ERR1218543	Urine	Thailand
Runchaeron 2017	3898STDY6199615	ERR1218615	Blood	Thailand
Runchaeron 2017	3898STDY6199616	ERR1218616	Urine	Thailand
Runchaeron 2017	3898STDY6199617	ERR1218617	Blood	Thailand
Runchaeron 2017	3898STDY6199618	ERR1218618	Blood	Thailand
Runchaeron 2017	3898STDY6199619	ERR1218619	Urine	Thailand
Runchaeron 2017	3898STDY6199620	ERR1218620	Urine	Thailand
Runchaeron 2017	3898STDY6199621	ERR1218621	Blood	Thailand
Runchaeron 2017	3898STDY6199622	ERR1218544	Urine	Thailand
Runchaeron 2017	3898STDY6199623	ERR1218622	Urine	Thailand
Runchaeron 2017	3898STDY6199624	ERR1218545	Blood	Thailand
Runchaeron 2017	3898STDY6199625	ERR1218623	Urine	Thailand
Runchaeron 2017	3898STDY6199626	ERR1218624	Urine	Thailand
Runchaeron 2017	3898STDY6199627	ERR1218625	Urine	Thailand
Runchaeron 2017	3898STDY6199628	ERR1218626	Urine	Thailand
Runchaeron 2017	3898STDY6199629	ERR1218627	Urine	Thailand
Runchaeron 2017	3898STDY6199630	ERR1218628	Blood	Thailand
Runchaeron 2017	3898STDY6199631	ERR1218629	Blood	Thailand
Runchaeron 2017	3898STDY6199632	ERR1218630	Pus	Thailand
Runchaeron 2017	3898STDY6199633	ERR1218631	Pus	Thailand
Runchaeron 2017	3898STDY6199634	ERR1218632	Urine	Thailand
Runchaeron 2017	3898STDY6199635	ERR1218633	Blood	Thailand
Runchaeron 2017	3898STDY6199636	ERR1218634	Pus	Thailand
Runchaeron 2017	3898STDY6199637	ERR1218635	Urine	Thailand
Runchaeron 2017	3898STDY6199638	ERR1218546	Urine	Thailand
Runchaeron 2017	3898STDY6199639	ERR1218636	Pus	Thailand
Runchaeron 2017	3898STDY6199640	ERR1218637	Urine	Thailand
Runchaeron 2017	3898STDY6199642	ERR1218639	Blood	Thailand
Runchaeron 2017	3898STDY6199643	ERR1218640	Pus	Thailand
Runchaeron 2017	3898STDY6199644	ERR1218641	Blood	Thailand
Runchaeron 2017	3898STDY6199645	ERR1218642	Canal	Thailand
Runchaeron 2017	3898STDY6199648	ERR1218643	Canal	Thailand

Table 6.3: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Runchaeron 2017	3898STDY6199649	ERR1218644	Canal	Thailand
Runchaeron 2017	3898STDY6199651	ERR1218549	Canal	Thailand
Runchaeron 2017	3898STDY6199653	ERR1218551	Canal	Thailand
Runchaeron 2017	3898STDY6199654	ERR1218646	Canal	Thailand
Runchaeron 2017	3898STDY6199656	ERR1218552	Canal	Thailand
Runchaeron 2017	3898STDY6199657	ERR1218648	Canal	Thailand
Runchaeron 2017	3898STDY6199658	ERR1218553	Canal	Thailand
Runchaeron 2017	3898STDY6199659	ERR1218649	Canal	Thailand
Runchaeron 2017	3898STDY6199660	ERR1218650	Canal	Thailand
Runchaeron 2017	3898STDY6199661	ERR1218651	Canal	Thailand
Runchaeron 2017	3898STDY6199662	ERR1218652	Canal	Thailand
Runchaeron 2017	3898STDY6199664	ERR1218654	Canal	Thailand
Runchaeron 2017	3898STDY6199665	ERR1218655	Canal	Thailand
Runchaeron 2017	3898STDY6199667	ERR1218656	Canal	Thailand
Runchaeron 2017	3898STDY6199669	ERR1218658	Canal	Thailand
Runchaeron 2017	3898STDY6199670	ERR1218659	Canal	Thailand
Runchaeron 2017	3898STDY6199671	ERR1218660	Canal	Thailand
Runchaeron 2017	3898STDY6199672	ERR1218661	Canal	Thailand
Runchaeron 2017	3898STDY6199673	ERR1218662	Canal	Thailand
Runchaeron 2017	3898STDY6199674	ERR1218663	Canal	Thailand
Runchaeron 2017	3898STDY6199675	ERR1218664	Canal	Thailand
Runchaeron 2017	3898STDY6199677	ERR1218666	Canal	Thailand
Runchaeron 2017	3898STDY6199680	ERR1218669	Canal	Thailand
Runchaeron 2017	3898STDY6199682	ERR1218671	Canal	Thailand
Runchaeron 2017	3898STDY6199685	ERR1218674	Canal	Thailand
Runchaeron 2017	3898STDY6199686	ERR1218675	Canal	Thailand
Runchaeron 2017	3898STDY6199687	ERR1218676	Canal	Thailand
Runchaeron 2017	3898STDY6199689	ERR1218678	Canal	Thailand
Runchaeron 2017	3898STDY6199692	ERR1218681	Canal	Thailand
Runchaeron 2017	3898STDY6199693	ERR1218682	Canal	Thailand
Runchaeron 2017	3898STDY6199694	ERR1218683	Canal	Thailand
Runchaeron 2017	3898STDY6199695	ERR1218684	Canal	Thailand
Runchaeron 2017	3898STDY6199696	ERR1218685	Canal	Thailand
Runchaeron 2017	3898STDY6199697	ERR1218686	Canal	Thailand
Runchaeron 2017	3898STDY6199698	ERR1218554	Canal	Thailand
Runchaeron 2017	3898STDY6199700	ERR1218688	Canal	Thailand
Runchaeron 2017	3898STDY6199701	ERR1218689	Canal	Thailand
Runchaeron 2017	3898STDY6199702	ERR1218690	Untreated hospital sewage	Thailand
Runchaeron 2017	3898STDY6199704	ERR1218692	Untreated hospital sewage	Thailand
Runchaeron 2017	3898STDY6199705	ERR1218693	Untreated hospital sewage	Thailand
Runchaeron 2017	3898STDY6199706	ERR1218694	Canal	Thailand
Runchaeron 2017	3898STDY6199707	ERR1218695	Canal	Thailand
Runchaeron 2017	3898STDY6199708	ERR1218696	Canal	Thailand
Runchaeron 2017	3898STDY6199709	ERR1218697	Farm	Thailand
Runchaeron 2017	3898STDY6199710	ERR1218698	Farm	Thailand
Runchaeron 2017	3898STDY6199713	ERR1218701	Farm	Thailand
Runchaeron 2017	3898STDY6199714	ERR1218702	Farm	Thailand
Runchaeron 2017	3898STDY6199715	ERR1218703	Farm	Thailand
Runchaeron 2017	3898STDY6199764	ERR1218705	Urine	Thailand
Runchaeron 2017	3898STDY6199766	ERR1218556	Pus	Thailand
Runchaeron 2017	3898STDY6199768	ERR1218706	Urine	Thailand
Runchaeron 2017	3898STDY6199769	ERR1218707	Urine	Thailand
Runchaeron 2017	3898STDY6199772	ERR1218708	Pus	Thailand
Runchaeron 2017	3898STDY6199773	ERR1218557	Pus	Thailand
Runchaeron 2017	3898STDY6199778	ERR1218709	Canal	Thailand
Runchaeron 2017	3898STDY6199780	ERR1218710	Canal	Thailand
Runchaeron 2017	3898STDY6199781	ERR1218558	Canal	Thailand
Runchaeron 2017	3898STDY6199784	ERR1218711	Canal	Thailand
Runchaeron 2017	3898STDY6199790	ERR1218559	Urine	Thailand
Runchaeron 2017	3898STDY6199792	ERR1218712	Blood	Thailand
Runchaeron 2017	3898STDY6199793	ERR1218713	Urine	Thailand
Runchaeron 2017	3898STDY6199796	ERR1218560	Pus	Thailand

Table 6.3: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Runchaeron 2017	3898STDY6199798	ERR1218714	Blood	Thailand
Runchaeron 2017	3898STDY6199799	ERR1218715	Canal	Thailand
Runchaeron 2017	3898STDY6199802	ERR1218716	Canal	Thailand
Runchaeron 2017	3898STDY6199804	ERR1218561	Canal	Thailand
Runchaeron 2017	3898STDY6199805	ERR1218717	Canal	Thailand
Runchaeron 2017	3898STDY6199806	ERR1218718	Canal	Thailand
Runchaeron 2017	3898STDY6199807	ERR1218719	Canal	Thailand
Runchaeron 2017	3898STDY6199808	ERR1218562	Canal	Thailand
Runchaeron 2017	3898STDY6199809	ERR1218720	Canal	Thailand
Runchaeron 2017	3898STDY6199815	ERR1218564	Farm	Thailand
Runchaeron 2017	3898STDY6199816	ERR1218723	Canal	Thailand
Runchaeron 2017	3898STDY6199923	ERR1218774	Canal	Thailand

*Note:*

Accession numbers for this study starting ERS are not yet published on GenBank.

RS = Rectal swab

## Chapter 7

# Whole genome sequencing to track longitudinal ESBL-E colonisation

### 7.1 Chapter overview

In this chapter, I present an attempt to use short read whole genome sequence (WGS) data as a high resolution typing tool to track bacteria and mobile genetic elements (MGE) within the participants of the study. I use the hierarchical BAPS (Bayesian analysis of population structure) algorithm to classify the core gene alignment of 473 *E. coli* into 48 sequence clusters, which mapped well to the phylogeny. I used the *cd-hit* algorithm to cluster 488 ESBL-gene containing contigs into 99 clusters, which showed some lineage association on mapping them back to the phylogeny. Largely, hospital associated isolates were independent of sequence cluster assignment with the exception of sequence cluster 23 which was associated with hospital acquisition ( $p = 6.3 \times 10^{-4}$ ), and corresponded to the putative recently arrived high-risk clone ST410 described in Chapter 6.

The combination of hierBAPS sequence cluster and ESBL-contig cluster together were conserved within participants on longitudinal sampling compared to between-participants ( $p = 1.1 \times 10^{-12}$ ) whereas either alone was not ( $p = 0.4$  and  $p = 1.0$ ). However beyond 35 days apart any two samples from a single a participant were only as likely to contain the same sequence cluster-ESBL cluster combination as any two samples randomly selected from the dataset. This suggests that, firstly, the unit of transmission in this system is likely to be the bacterium rather than the MGE. Secondly, the within-participant ESBL contig cluster bacterial sequence cluster association suggests that a given bacterium-MGE combination is reasonably stable on the timescale of the study. Finally, these data suggest there is turnover of ESBL *E. coli* on a time scale of around 35 days suggestive of either frequent re-exposure or some other endogenous turnover in the microbiota.

## 7.2 Introduction and chapter aims

In Chapter 5, I described the epidemiology of ESBL-E carriage in study participants as they were exposed to antimicrobials and hospitalisation, showing a dramatic increase in carriage prevalence following hospitalisation and, particularly, hospitalisation and antibacterial exposure. In Chapter 6, I presented details of the whole genome sequencing of a sample of 473 *E. coli* isolates recovered from the stool of these study participants, placed them in a global context, and described the antimicrobial resistance (AMR) determinants that the isolates carried. In this chapter, I present an analysis whereby I combine the WGS data with the metadata from Chapter 5 in an attempt to use WGS as a high-resolution typing tool to track carriage of ESBL *E. coli* within participants in this study. Specifically, the aims of this chapter are:

1. To use clustering algorithms to classify clusters of bacteria and MGE that can be used in further analyses to track bacteria and MGE within participants and associations between bacteria and MGE and metadata.
2. To explore whether apparent hospital acquisitions of ESBL *E. coli* can be distinguished on a genomic level from community associated *E. coli*.
3. To use bacteria and MGE clusters to determine a), which, if any, element is conserved within participants over time and b) the time scale over which bacteria and MGE change over time within-host.

## 7.3 Methods

The collection, culture, whole genome sequencing and bioinformatic analysis of the isolates analysed in this chapter are described in Chapter 2, Methods, and Chapter 7. The methods of the further analyses covered in this chapter are described here.

The *rhierBAPS* v1.1.0 package in R[370] was used to cluster the core genome pseudosequence into sequence clusters (SCs). Two levels were used and these level 2 clusters used to test associations (see statistical analysis, below). To track putative mobile genetic elements ESBL-gene containing contigs were identified using BLASTn v2.7.0[371] of all contigs against the SRST2 database[343] of AMR genes (the same database used with ARIBA in the previous chapter). Contigs containing any given ESBL gene were grouped by the ESBL gene they contained (for example, all *bla<sub>CTXM-15</sub>* gene-containing clusters were grouped together). Each group was then clustered using cd-hit v4.6[372] to produce mutually exclusive ESBL-gene-containing contig clusters for each identified ESBL gene. Henceforth, these clusters will be referred to as ESBL-clusters, for brevity.



In order to attempt to determine the biological significance of the identified ESBL-clusters (i.e. what kind of MGE element they are likely to represent), basic statistics were plotted: number of samples contained within each cluster, length of longest contig in cluster in kbases, length distribution of all contigs in cluster relative to longest contig and distribution of sequence identity compared to the longest contig in the cluster. Presence of insertion sequences (i.e compound transposons), AMR determinants and plasmid replicons were identified for ESBL-cluster representative sequence (as determined by cd-hit i.e one, the longest, for each ESBL-cluster) using BLAST. BLAST default settings were used against the insertion sequence finder (ISfinder) database[373] and the SRST2 database, filtering such that sequence identity was greater or equal to 95%, taking the top hit (as determined by bitscore) for any given location if there were two overlapping hits, and visualising the results in *ggenes* v0.3.2 package in R. To assess lineage association, the ESBL-clusters were mapped back to the core gene phylogeny.

To explore hospital or community associations of isolated *E. coli*, the location of isolation was first mapped onto the phylogenetic tree; location of isolation was classified as hospital, community, or recent hospital discharge (defined as a date of isolation within 2 weeks of hospital discharge). This latter category was used because it is possible that a patient could acquire an ESBL-E clone in hospital but only be sampled once leaving hospital; using only hospital isolated and community isolated categories could therefore introduce bias. Hospital or community association of each sequence cluster was assessed using a Fisher's exact test of proportion of hospital associated samples (defined as sum of hospital isolated and recent hospital discharge) for the given sequence cluster as compared to proportion of hospital associated samples in the remainder of the samples, with a Bonferroni correction for multiple comparisons.  $p < 0.05$  was again considered statistically significant.

To compare within-patient to between-patient conservation of bacteria (as represented by sequence cluster) and ESBL-containing MGE (as represented by the ESBL-clusters) several approaches were taken. Firstly, I assessed whether either sequence cluster or ESBL-cluster were conserved within an individual at all. I hypothesised that within-patient correlation is a function of time: samples closer together in time are more likely to be similar. To assess if this was the case for bacteria, pairwise core genome pseudosequence SNP distance was calculated using *snp-dists* v0.4 (<https://github.com/tseemann/snp-dists>) for all samples and plotted against the time difference (in days) between samples, within and between patients, and with a smoothed curve fitted using a general additive model (GAM) with cubic splines. Because of significant overplotting, this was also plotted as a 2D density plot. Based on these plots, the within and between patient SNP distances were compared in two post-hoc defined groups binned by time distance between the samples (50 days or less vs. more than 50 days, cutoffs that were determined from inspection of the pairwise SNP distance vs time plots), and distributions compared with Kruskal-Wallis tests.

I then compared the within patient temporal clustering of ESBL-clusters and sequence clusters, by estimating the proportion of within-patient samples that contain the same ESBL-cluster or sequence cluster, as a function of time; essentially a temporal auto-correlation function. To estimate this, I considered pairwise comparison of all within-patient samples. For any given time ( $t$ ) between samples I defined a window of  $\pm 5$  days and estimated the probabilities as the number of all within-patient sample pairs in the window  $[t - 5, t + 5]$  that contained the same sequence cluster or ESBL-cluster divided by the total number of all within-patient sample pairs within that time window. Exact binomial confidence intervals for these proportions were generated and probabilities plotted as a function of time. In order to estimate the probability of two samples containing the same sequence cluster or contig-cluster purely by chance, 1000 sample pairs were randomly drawn from all samples with replacement and the proportion of these samples that contained the same sequence cluster or ESBL-cluster calculated.

Finally, to inform the question as to what the likely unit of transmission in this system is, I assessed what was most conserved within patients, in pairwise sample comparison: bacteria (as represented by core gene sequence cluster), ESBL-containing MGE (as represented by ESBL-cluster), or both. Simple proportions in all-against-all pairwise comparison - stratified by whether between-patient or within-patient - were calculated: the proportion of samples that contain the same core gene sequence cluster only, the proportion of samples that contain the same ESBL-cluster only, and the proportion that contain both sequence cluster and ESBL-cluster. Proportions were compared between within and between-patient strata in these three groups using Fisher's exact test, with  $p < 0.05$  considered statistically significant.

## 7.4 Results

As described above, in order to test metadata associations of bacterial lineages or MGE, I used several techniques: considering core gene SNP distance between isolates to infer continuous carriage and/or transmission events, and clustering core gene pseudosequences and ESBL-containing contigs into mutually exclusive groups which can then be used to test associations. Below, I first describe the outcomes of the clustering algorithms used, before describing tests of association of the results with metadata.

### 7.4.1 Hierarchical BAPS clustering of core gene pseudosequences

The hierarchical BAPS algorithm clustered the core gene alignments into 15 level one (top level) clusters, denoted sequence clusters A-O, and a total of 48 level two (lower level) clusters, denoted sequence clusters 1-48 that were almost exclusively monophyletic and

often corresponded closely to the multilocus sequence types (STs, Figure 7.1A). Intracluster pairwise SNP distance varied (Figure 7.1B) but the clusters were often reasonably clonal: SC6, SC8 and SC23, for example (the three largest clusters) had median (IQR) intragroup pairwise SNP distance of 62 (34-97), 326 (18-378) and 18 (11-24) respectively.

#### 7.4.2 ESBL-clusters

The 473 samples contained 486 ESBL genes (Figure 7.2A); 5 genes only occurred once in the collection and so no attempt was made to cluster them. Of the remaining 481 genes, BLAST failed to identify the ESBL-gene containing contig in 2 samples (one in which ARIBA had identified *bla<sub>CTXM-15</sub>* one *bla<sub>CTXM-27</sub>*), but identified the remaining 479 ESBL genes on 478 contigs, with perfect agreement with ARIBA as to which AMR gene was present in which sample. Only one contig carried two ESBL genes: *bla<sub>CTXM-3</sub>* and *bla<sub>CTXM-15</sub>*; the remaining 477 contigs contained one. The *cd-hit* algorithm grouped the 477 unique contigs into 99 clusters (Figure 7.2B). In total, over 90% of the ESBL-genes (432/479 [90%]) were contained in the 52 largest contig clusters.

The *cd-hit* algorithm selects one member of a cluster (the longest) as the representative. The structure of these representative contigs was explored in an attempt to understand the type of MGE they were likely to represent. The length of the representative clusters was very variable, ranging from 1.8kbp to 905.8kbp, with median (IQR) 46.1kbp (11.1-215.5kbp). The other cluster members were usually fragments of these representative contigs with varying sizes - a median (IQR) 60% (36-100%) of the representative contig length - but had high sequence identity, median (IQR) 100.0% (99.7-100.0%) (Figure 7.6 in the appendix to this chapter).

I then explored the insertion sequence (IS), AMR gene and plasmid replicon content of the representative contig for each cluster using BLAST against the SRST2, ISfinder and Plasmidfinder databases (Figures 7.8 and 7.9 in the appendix to this chapter). Every ESBL gene was closely associated with at least one IS, commonly *ISEcp1*, *IS26* and *IS903B*. *IS26* was frequently associated with an apparent 108bp fragment of a *catB4* chloramphenicol resistance determinant. Some ESBL-genes were associated with particular IS; *bla<sub>CTXM-15</sub>*, *bla<sub>CTXM-9</sub>* and *bla<sub>CTXM-1</sub>*, for example were very commonly associated with *ISEcp1*, whereas *bla<sub>SHV-12</sub>* was associated with *IS26*. ESBL genes were not infrequently associated with other resistance determinants, including commonly *bla<sub>CTXM-15</sub>* with *bla<sub>TEM-1</sub>*. Plasmid replicons were occasionally identified, including an IncFIB plasmid carrying *bla<sub>CTXM-15</sub>* and an IncQ1 plasmid carrying *bla<sub>CTXM-27</sub>*. It is clear that the same configuration of AMR genes and IS are seen across different contigs, despite a varying backbone, implying historical transposition events. Finally, to assess lineage associations of the identified ESBL-clusters, I mapped the clusters back to the tree, and found that there

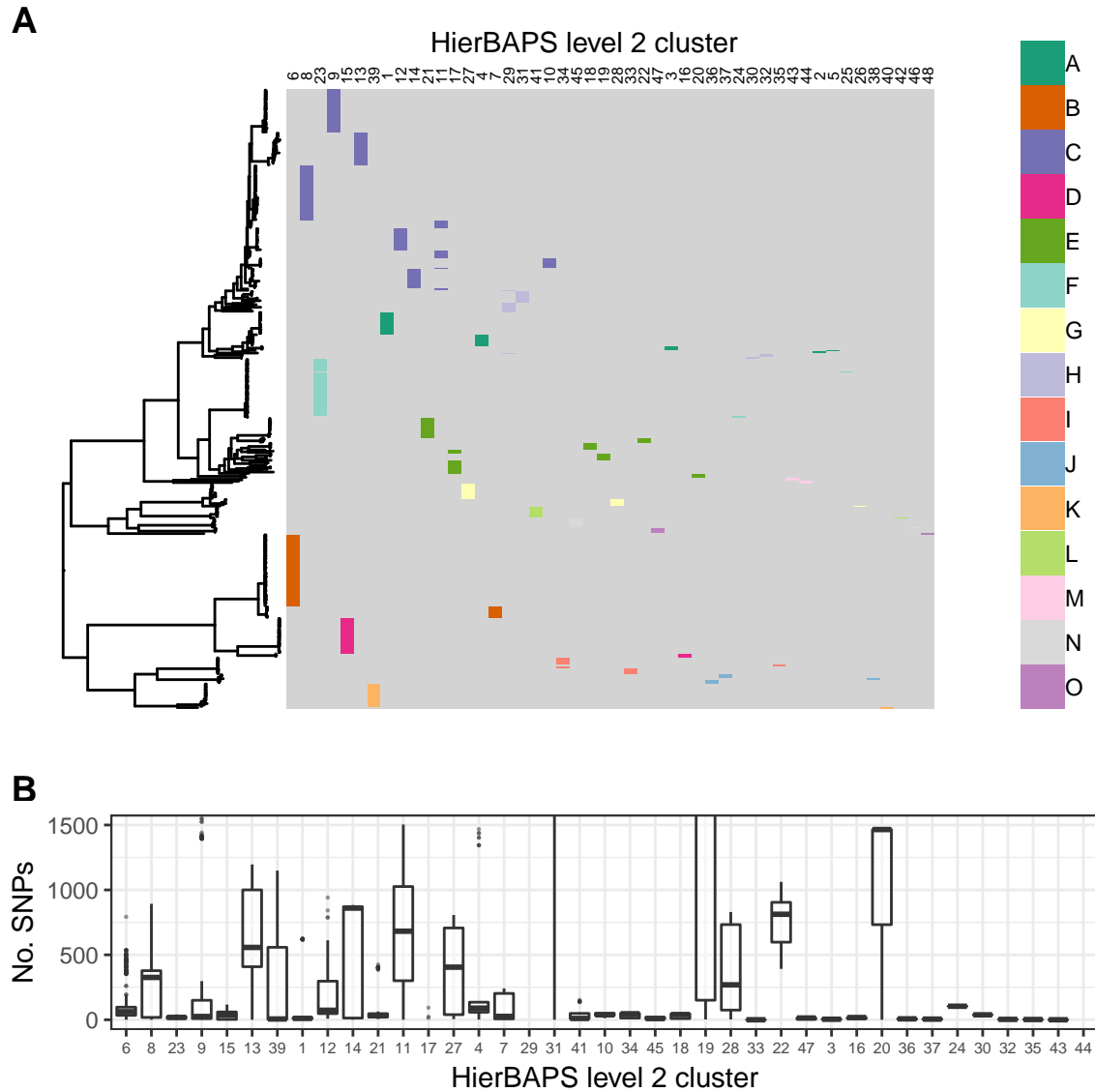


Figure 7.1: A: Core gene hierarchical BAPS clusters mapped back to phylogeny. Heatmap shows level 2 (lower level) with colour denoting level 1 (top level) cluster membership. B: Intracluster pairwise SNP distance for level 2 sequence clusters. Axis restricted to 0-1500 SNPs and as result SC17 (median 6881 SNPs), SC29 (median 2970 SNPs), SC31 (median 2970 SNPs) and SC44 (median 12322 SNPs) boxes are not shown. Boxplots show median and IQR, whiskers show 1.5 times IQR, and outliers are points falling beyond whiskers.

was often a strong lineage association (Figure 7.2C).

### 7.4.3 Assessing for healthcare-associated lineages

Having clustered bacteria and MGE using *rhierBAPS* and *cd-hit* respectively, I then mapped the location of sample collection back to the phylogeny and used the *hierBAPS* SCs to assess for healthcare associated lineages (Figure 7.3). In general, healthcare-associated isolates were distributed throughout the tree and across all SCs, rather than there being a clear hospital-associated lineage. The exception to this was SC23, corresponding to MLST 410, which was more likely to be healthcare associated. When comparing the proportion of healthcare associated samples within each SC to the remained of samples, only SC23 had a statistically significantly increased proportion of healthcare associated samples on Fisher's exact test ( $p = 6.3 \times 10^{-4}$ , threshold of significance following Bonferroni correction  $1.0 \times 10^{-3}$ ), though it was by no means health-facility restricted: 50% (21/42) of SC23 samples were isolated in the community.

### 7.4.4 Assessing for within-patient conservation of lineage or MGE

To answer the question as to what elements (bacteria or MGE) are conserved within individuals across time I first compared all-against-all pairwise SNP distance between and within patients; first as a scatter plot, and then, because of significant overplotting, as a density plot (Figure 7.4). This suggested that there are a cluster of points close to the origin in the within-patient plot that are not seen in the between-patient plot: before approximately 50 days, there are more similar within-patient isolates than seen in the between-patient isolates. Dichotomising time at 50 days (based on inspection of the density plots) and performing a Kruskal-Wallis test enabled me to identify a statistically significant difference between the before 50 day and after 50 day pairwise SNP distance distribution in the within patient stratum ( $p = 0.008$ ) but not in the between-patient stratum ( $p = 0.07$ ). After 50 days, the distribution of between- and within- patient SNP distances are similar ( $p = 0.45$ ). However it is clear from the plots that even at small  $t$  and within-participant, there is significant diversity in the SNP distances, and that some isolates close together in time, within the same participant, are only distantly related.

Having confirmed that there is a signal for within-participant temporal conservation of ESBL-E, I then sought to determine if the sequence clusters and ESBL-clusters were similarly conserved over time, and if so, which was the more conserved. The proportion of within-patient sample pairs that contained the same ESBL-cluster and sequence cluster were significantly greater than would be expected by chance when the time between the samples is less than 35 days for sequence cluster and 32 days for ESBL-cluster (Figure 7.5A). After this time, the lower confidence interval of the sequence cluster and ESBL-cluster

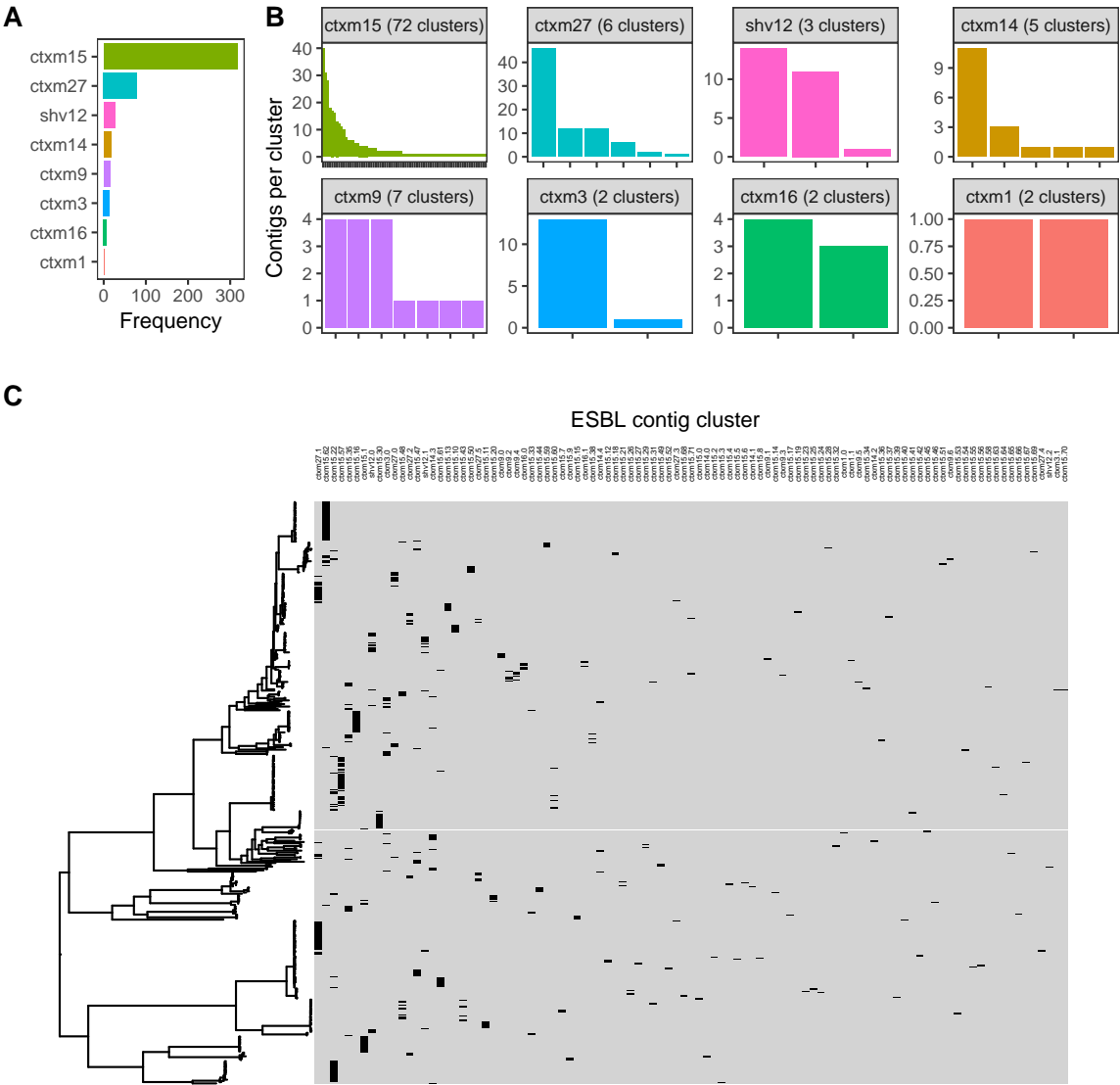


Figure 7.2: A: Frequency distribution of ESBL genes in included samples. B: Frequency distribution of samples per ESBL-cluster, stratified by gene. C: ESBL-cluster membership mapped back to phylogeny.

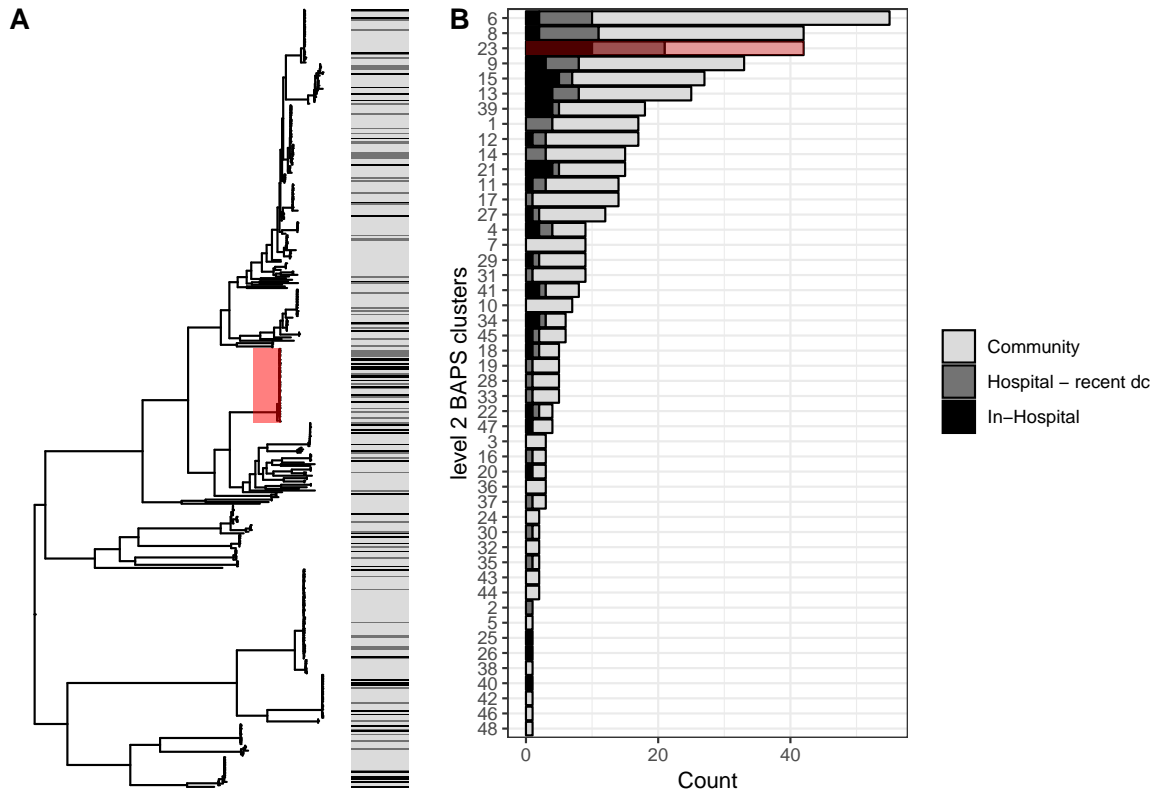


Figure 7.3: A: Location of sample isolation mapped back to phylogeny B: Distribution of location of sample isolation stratified by hierBAPS cluster. In each case, community isolates include those cultured from samples collected on the first day of hospital admission, in-hospital isolates are from patients who have been hospitalised > 24 hrs and recent discharge isolates are from patients who have been discharged from hospital within the last 2 weeks. Sequence cluster 23, highlighted in red, showed a statistically significant association with hospitalisation (see text).

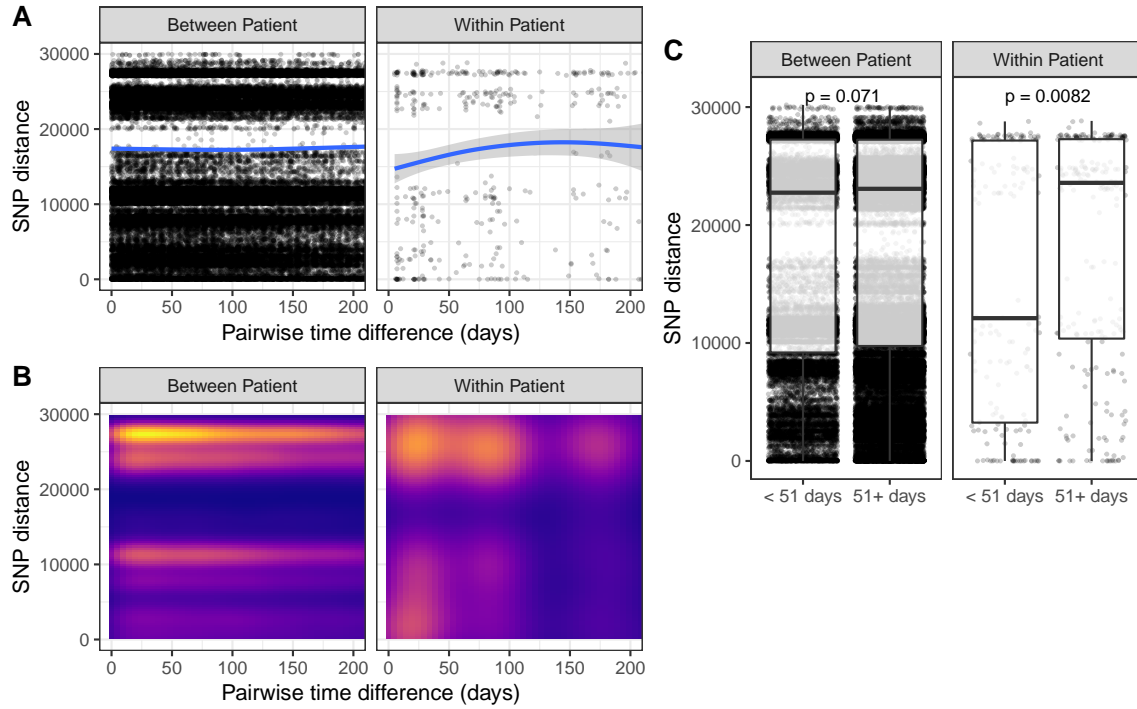


Figure 7.4: Within and between participant pairwise SNP distances. A: Scatterplot of pairwise SNP distances as a function of time with GAM model fitted curve. B: Pairwise SNP distance as function of time as a 2D density plot, showing cluster of isolates close to origin that are close together in time and SNP-distance. C: Pairwise SNP distance distribution before and after 50 days, within and between patients, showing statistically significant decrease in pairwise SNP distance within patients before 50 days. After 50 days, between- and within- patient distributions are similar.

curve crossed the proportion of samples that would be expected to be the same by chance, suggesting that, after 35 or 32 days, the chance of any two within-patient samples having the same sequence cluster or ESBL-cluster (respectively) is the same as if the two samples were randomly picked from the data set without regard to patient. The two curves have a very similar appearance; to address the question of which element is most conserved within an individual - sequence cluster, ESBL-cluster, or both - I performed an all-against-all pairwise comparison of which elements were conserved (Figure 7.5C), and found that only ESBL-cluster and sequence cluster together are conserved within patients at a significantly greater proportion than between patients ( $p = 1.1 \times 10^{-12}$ ).

## 7.5 Discussion

In this chapter, I have used clustering algorithms on WGS data to group bacteria and putative MGE. I have then assessed associations of these groups with metadata to attempt to better understand the determinants of ESBL-E carriage in Blantyre. I draw several conclusions.



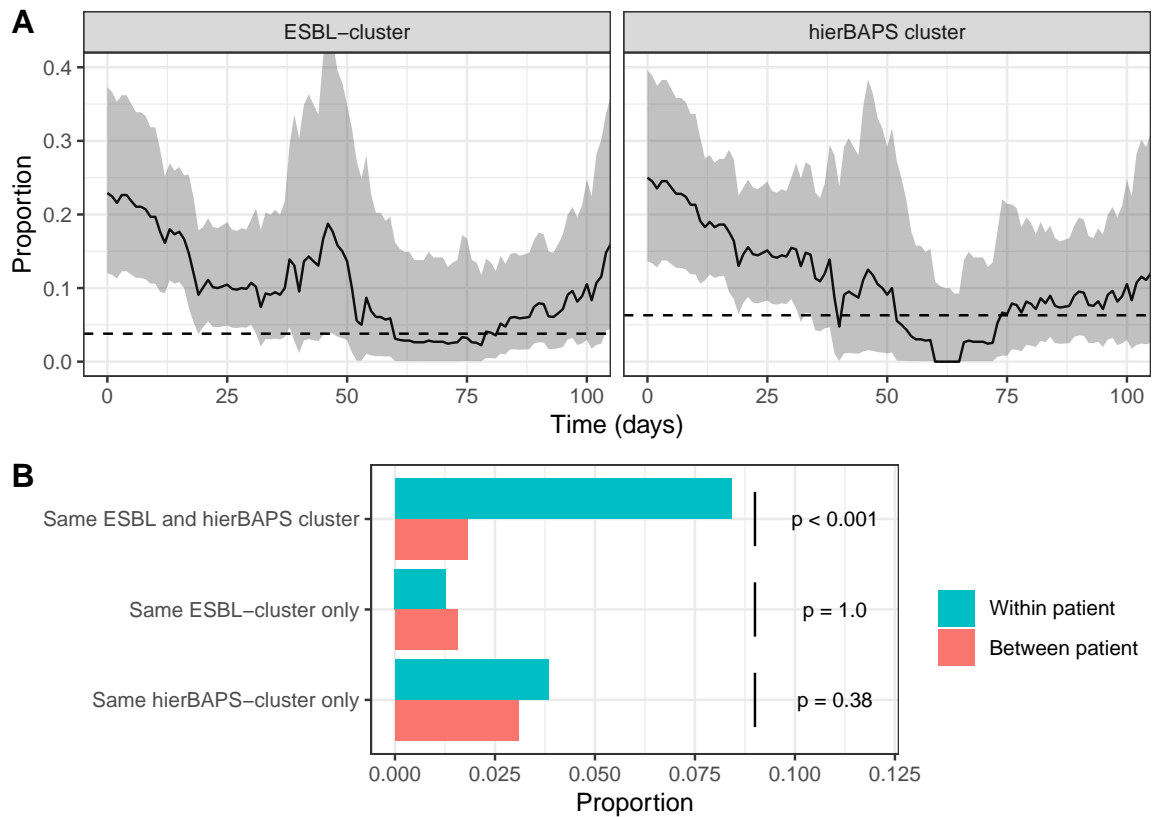


Figure 7.5: Probability of any two samples from within a given participant containing the same ESBL-cluster (A, left panel) or being a member of the same hierBAPS cluster (A, right panel). Time is windowed at  $\pm 5$  days around the time indicated on the x axis. Dotted line is the probability that two samples would belong to the same group by chance, constructed by randomly sampling 1000 sample pairs. B: proportion of samples that contain the same hierBAPS cluster alone, or ESBL-cluster alone, or both, demonstrating that the ESBL cluster-hierBAPS cluster pairing is the most conserved of the three.

First, I looked for healthcare associated *E. coli* sequence clusters. Healthcare associated bacteria were not associated with a particular sequence cluster, and were spread throughout the phylogeny rather than apparent hospital acquisitions being restricted to a single clade or clone. The exception to this was SC23, which corresponds to ST410, and was more likely to be healthcare associated. This could be consistent with the hypothesis that it is a recently arrived high-risk clone, which may be, at least initially, hospital-associated. Even so, it is clearly not hospital restricted, with half of the ST410 isolates being isolated from the community.

I showed in Chapter 5 that there was an increase in colonisation prevalence of ESBL-E in study participants following admission to hospital, particularly in the antibiotic-exposed. Despite this it seems as though the genomic diversity of ESBL *E. coli* apparently acquired in hospital is largely the same as *E. coli* isolated in the stool of community members. This result could be explained by two hypotheses: first, that these are true transmission events that are occurring within the hospital, and that the diversity of ESBL *E. coli* within the hospital is the same as the community; or, second, that these “hospital acquisitions” are actually minority variant *E. coli* present in the microbiota (and therefore acquired in the community) at a low abundance and hence not detected by culture, and enriched for with antimicrobial exposure in hospital. Distinguishing between these two hypotheses is important as they would each require a different intervention: hospital infection control in the former case, or strategies to protect the microbiota from the deleterious effect of broad spectrum antimicrobials (such as pre- or probiotics, or oral  $\beta$ -lactamases) in the latter. It is of course possible that both hypotheses are true; they are not mutually exclusive. The genomic data are consistent with both, perhaps with a suggestion from the hospital association of ST410 that there is at least some true hospital acquisition. The longitudinal models I present in the next chapter can help to shed more light on the mechanism of increase of ESBL-E prevalence following admission, by quantifying the relative effects of hospital admission versus antimicrobial exposure.

By forming sequence clusters and ESBL-clusters, I was able to demonstrate that bacteria and MGE are conserved together, within-patient, over time, whereas bacteria and MGE alone are not. Some previous longitudinal studies of ESBL-E found that *E. coli* STs tended to vary over time but that in many cases ESBL gene and plasmid replicons remained the same, which could be due to a conserved MGE transferring between bacteria[177]. Given my findings, this is unlikely to be the case. Though not directly addressed in this study it is possible to speculate therefore that the unit of transmission of ESBL between patients is likely to be the bacterium rather than, for example, horizontal gene transfer of ESBL genes on plasmids or transposons. The within-participant association of sequence cluster with ESBL-cluster suggests that MGE are reasonably conserved within bacteria, at least on the timescale of the study. Mapping the ESBL-clusters back to the *E. coli* phylogeny also shows some lineage association, which is consistent with this. The within-patient

correlation of SC and ESBL-cluster lasts only for 32-35 days; two samples from a single patient more than 35 days apart are as likely to contain the same SC/ESBL cluster as two samples from two different patients. This implies either an exogenous re-exposure or some other endogenous mechanism whereby the dominant ESBL strain is replaced by a minority variant from within the microbiota. Again, the longitudinal models in the next chapter will investigate this further.

This analysis also is suggestive that there is significant within-participant diversity of ESBL *E. coli*. At a maximum (i.e. with samples that are days apart), only around 20% of within-patient samples contain the same SC/ESBL-cluster. This is many times more than would be expected by chance, but still implies that at any time point there is significant diversity of ESBL *E. coli* strains, that have been missed by only taking forward one colony pick for sequencing. Without multiple picks from one time point, however, it is not possible to fully define within-host diversity. Limited data are available to define this diversity; one study that examined this question found that it varied between individuals, and that some individuals harboured widespread diversity of STs and ESBL genes whereas some did not[374].

### 7.5.1 Limitations

Only one colony pick from the ESBL selective media was taken forward for sequencing. In effect, we have randomly sampled one strain from all available strains at any give time point. This is likely to result in an underestimation of the extent to which strains persist within the individual over time, as strains that are present (but not sampled) are classed as absent in the above analysis.

Community members are under represented in this dataset; I have classified isolates as community or hospital associated, but there may have been healthcare contacts (especially in arm 1 and 2 of the study) which have not been recorded which would mean that isolates that were actually healthcare associated were classed as community associated. Healthcare contact is probably less likely in the true community arm of the study (arm 3).

I have attempted to overcome the difficulty of reconstructing MGE by defining ESBL-clusters as a proxy for MGE, but this approach has limitations. Some of the assembled contigs are short and likely represent transposons; the same transposons have likely inserted into multiple plasmids in the past and as such, these short contigs may cluster with other sequences that would be seen to be very different, were a full assembly available. In addition, the biological significance of these ESBL-clusters is not clear. It is not possible to say with certainty what they represent (e.g. plasmids) as they are only fragments. Nevertheless, the fact that I have seen within-patient associations of the ESBL-clusters lends some support to their use, as erroneous clustering would be expected to bias any associations towards to

null.

## 7.6 Conclusions and further work

In this chapter, I have shown that apparent hospital acquired ESBL *E. coli* are largely as diverse as community isolates. This suggests either widespread mixing of strains between the hospital and community and/or an enrichment effect as study participants are admitted to hospital and exposed to antimicrobials and carried but undetected ESBL *E. coli* become detectable by culture. By clustering bacteria and putative MGE I find that the bacteria-MGE combination is the element that is most conserved within-participant over time, but not after 32-35 days, suggestive of a constant re-exposure or some other replacement mechanism. Many questions remain unanswered and further work is planned:

- Shotgun metagenomic sequencing of stool would allow testing of the competing acquisition and unmasking hypotheses of rapid increase in ESBL-E prevalence by defining the microbiota and total AMR gene content pre-, during and post- antimicrobial exposure. This would also provide an opportunity to explore the role of the microbiota to colonisation resistance to ESBL-E, and help to define the within-host ESBL diversity at any time point.
- Long read sequencing would allow a proper characterisation of the MGE that carry ESBL genes in the Malawian context, giving the resolution necessary to truly track MGE within and between patients and strains.
- Short-read sequencing of the *Klebsiella pneumoniae* isolates from this study (as discussed in the previous chapter) would allow a comparison between the mechanisms of AMR and MGE prevalent in this species as compared to *E. coli*, and assess the extent to which horizontal gene transfer between the two is driving ESBL spread in Blantyre.
- Sequencing one *E. coli* from each sample in which *E. coli* was identified but has not yet had a representative sequenced (i.e. samples collected after October 2018) will give more power to detect metadata associations - in particular by expanding the number of isolates from true community members (arm 3) of the study, under represented in this dataset.
- Finally, incorporating the resolution afforded by sequencing into a longitudinal modelling approach may provide new insights in to the dynamics of ESBL-E carriage. This is taken up in the next chapter.

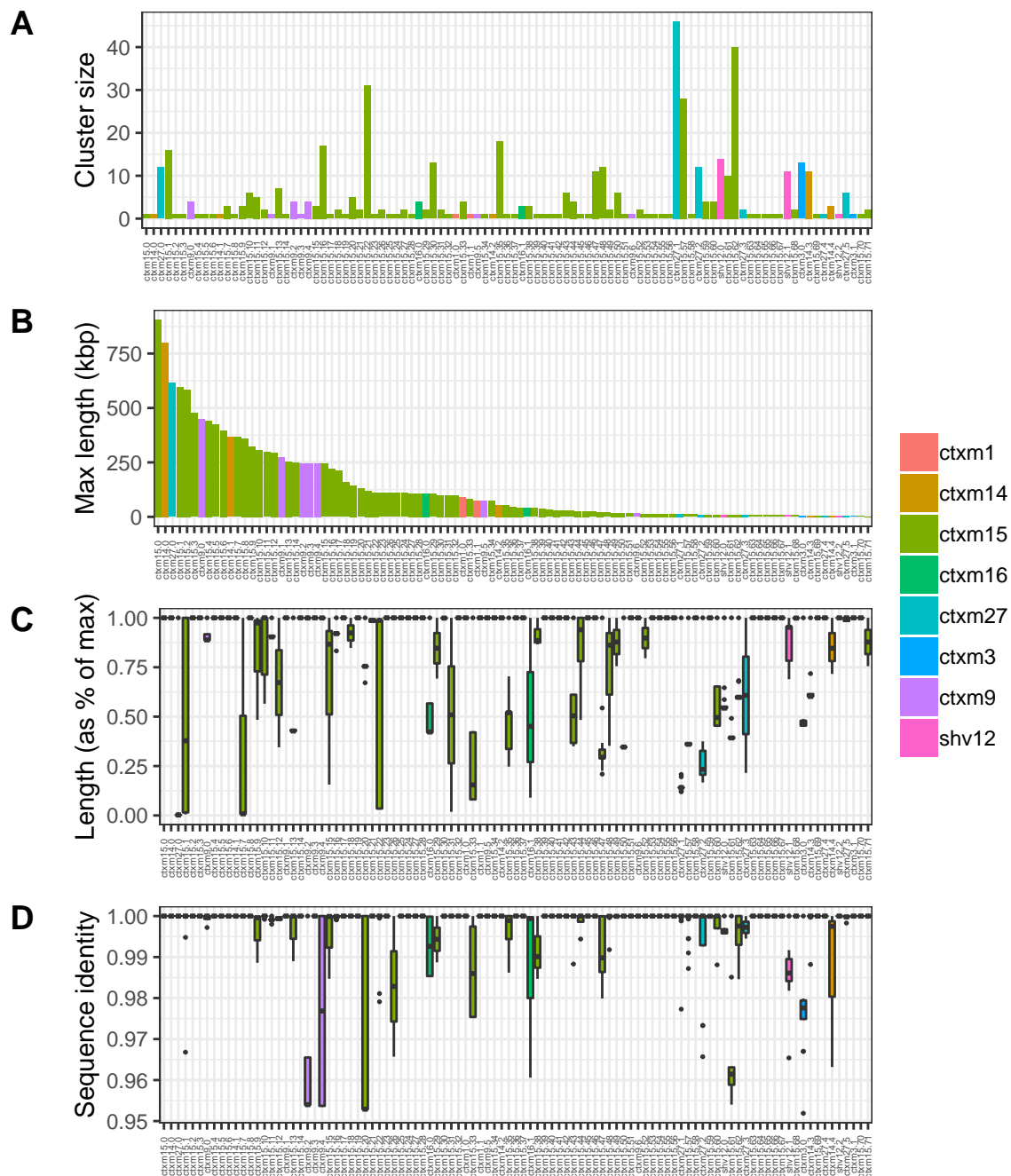


Figure 7.6: Summary statistics for 99 ESBL-containing contig clusters as determined by *cd-hit*. A: Number of contigs per cluster. B: Length (kbp) of longest sample in each cluster. This is defined as the cluster representative sample by *cd-hit* to which all other samples are compared for the purposes of length and sequence identity. C: Distribution of contig lengths by cluster expressed as a proportion of longest contig length. D: Distribution of sequence identity of cluster members compared to representative member, by cluster.



Figure 7.7: AMR genes, insertion sequences (IS) and plasmid replicons identified in the representative contig of each contig cluster, for *blaCTXM15* containing contigs, ordered by number of samples of cluster from most (top) to least (bottom). IS26 is very frequently associated with a 108bp fragment of *catB4* chloramphenicol resistance gene, shown as a red fragment within the green IS26 element. Plots show furthest IS/AMR gene or plasmid replicon up to  $\pm 10,000$ bp from the ESBL gene of interest.

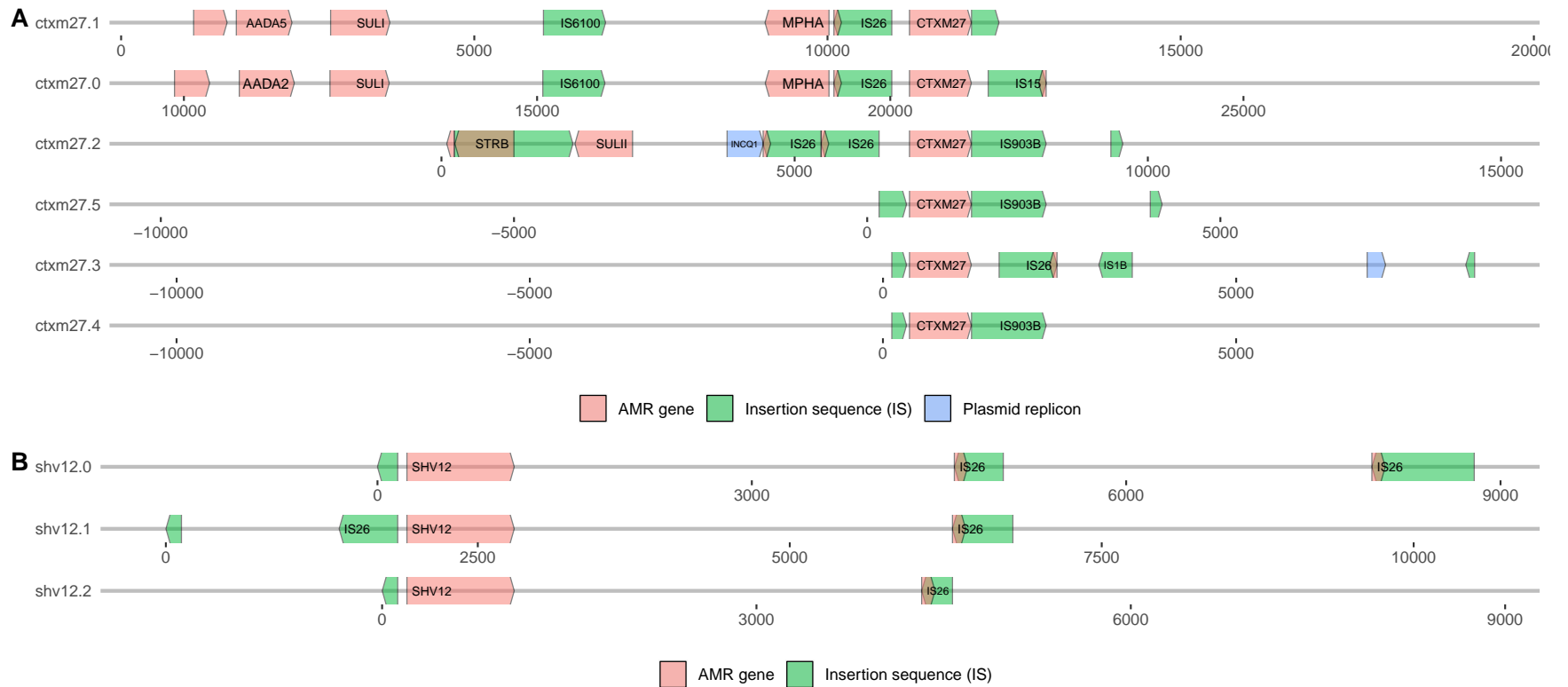


Figure 7.8: AMR genes, insertion sequences (IS) and plasmid replicons identified in the representative contig of each contig cluster for *blaCTXM27* (A) and *blaSHV12* (B). Ordered by number of samples of cluster from top (largest) to bottom (smallest) for each ESBL gene. IS26 is very frequently associated with a 108bp fragment of *catB4* chloramphenicol resistance gene, shown as a red fragment within the green IS26 element. Plots show furthest IS/AMR gene or plasmid replicon up to  $\pm 10,000$ bp from the ESBL gene of interest.

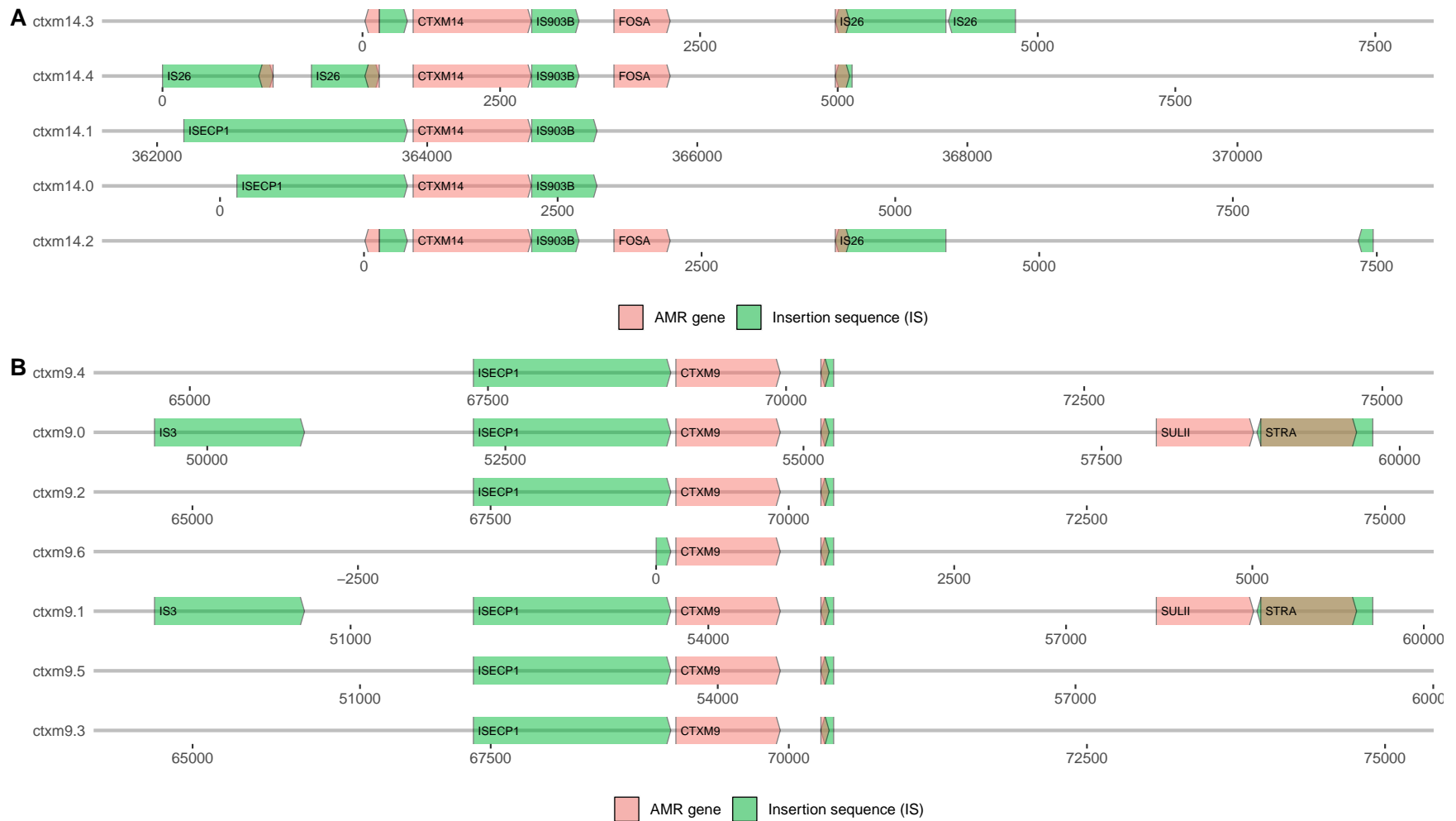


Figure 7.9: AMR genes, insertion sequences (IS) and plasmid replicons identified in the representative contig of each contig cluster for *blaCTXM14* (A) and *blaCTXM9* (B), ordered by number of samples in cluster from most (top) to least (bottom). IS26 is very frequently associated with a 108bp fragment of *catB4* chloramphenicol resistance gene, shown as a red fragment within the green IS26 element. Plots show furthest IS/AMR gene or plasmid replicon up to  $\pm 10,000$ bp from the ESBL gene of interest.



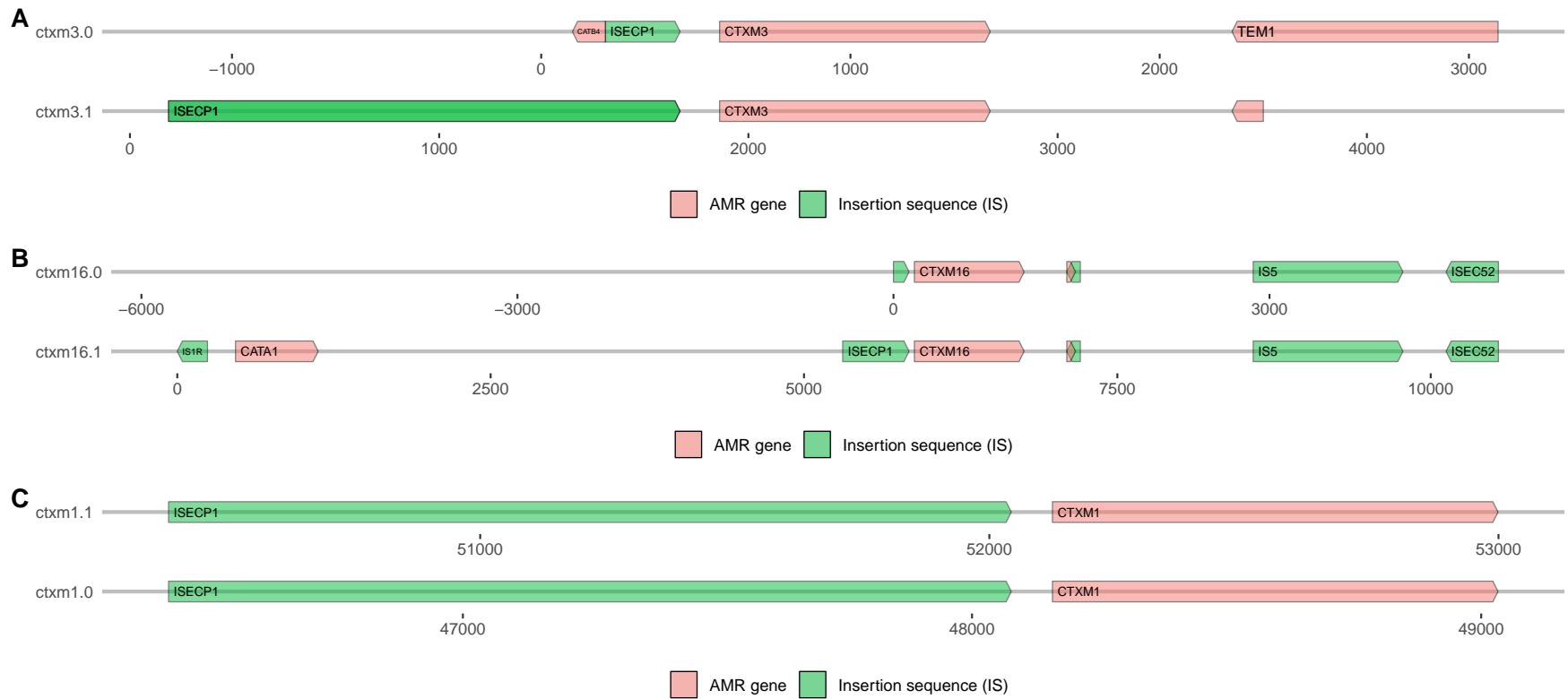


Figure 7.10: AMR genes, insertion sequences (IS) and plasmid replicons identified in the representative contig of each contig cluster for *bla*CTXM3 (A) and *bla*CTXM16 (B) and *bla*CTXM1 (C), ordered by number of samples in cluster from most (top) to least (bottom). IS26 is very frequently associated with a 108bp fragment of *catB4* chloramphenicol resistance gene, shown as a red fragment within the green IS26 element. Plots show furthest IS/AMR gene or plasmid replicon up to +/- 10,000bp from the ESBL gene of interest.



## Chapter 8

# Longitudinal Markov models of ESBL-E carriage

### 8.1 Chapter Overview

In this chapter, I develop time-inhomogeneous Markov models of ESBL-E carriage by individuals and fit them using the Bayesian probabilistic programming language Stan. I demonstrate that fitting these models is feasible with modest computational requirements, and that they are very flexible. I find that hospitalisation acts to increase both rate of ESBL-E acquisition and loss, with a net effect of rapidly increasing ESBL-E carriage prevalence; further, I find that antibacterial exposure acts to prolong ESBL-E carriage by reducing the rate of ESBL-E loss. It is, however, the synergistic effect of hospitalisation and antibacterial exposure that drives the rapid increase in ESBL-E carriage prevalence observed in antibacterial-exposed inpatients. I also find that co-trimoxazole preventative therapy (CPT) likely plays an important role as a determinant of long-term ESBL-E carriage. The models I have developed also support a post-exposure effect of antibacterials, such that they continue to have an effect long after they have been excreted from the body. I present hypotheses about the mechanism of such an effect along with the implications of my findings for antimicrobial stewardship interventions, and planned further work.

### 8.2 Introduction and chapter aims

In Chapter 5, I presented the longitudinal ESBL-E carriage data for the three arms of the clinical study that underpin this thesis. Antibacterial-exposed, hospitalised participants (arm 1) showed a rapid increase in ESBL-E carriage prevalence, whilst antibacterial-unexposed hospitalised participants (arm 2) showed a much more modest increase. This

suggests that antimicrobial exposure is the most significant determinant of acquisition of carriage; however, this unadjusted analysis is open to confounding. The participants recruited to the two arms of the study differed in important ways: antimicrobial unexposed participants were younger, less likely to be HIV-infected, with less cotrimoxazole preventative therapy (CPT) exposure, and crucially, a shorter length of hospital stay. An attempt to adjust for potential confounders using simple logistic regression models failed; therefore in this chapter I develop longitudinal models to quantify the relative roles of antibacterial exposure and hospitalisation in driving ESBL-E carriage.

There have been few prior attempts to model longitudinal ESBL-E carriage, and none where the focus was on the role of antimicrobials, including:

1. Fitting a Weibull distribution to community sample data[182]
2. Fitting a beta-distribution of admission and discharge carriage probability to data from trials of contact precautions in Dutch hospitals[375]
3. Modelling household ESBL-E acquisition as a Markov process[376].

None of these studies included the effect of antibacterial exposure as a covariate. The Markov model approach is an attractive method to model multi-state interval-censored data[377], has been used with a variety of clinical datasets[378–380] and is implemented in the *msm* package in R[381] where a maximum likelihood method is used to fit the models. However, *msm* allows only stepwise-constant covariate effects, largely for reasons of computational tractability in the maximum-likelihood framework, but there is no reason to assume that the effect of antibacterial exposure, for example, will act in this fashion. The aims of this chapter, therefore, are:

1. To generalise *msm*-type models to allow true time-varying covariates;
2. To demonstrate the feasibility of fitting such models;
3. To use the fitted models to infer an unbiased estimate of the relative roles of antibacterial exposure and hospitalisation in driving ESBL-E carriage by both fitted parameter estimates and simulating different levels of exposure;
4. To compare models with and without a post-exposure effect of antibacterials to assess the support in this data for such an effect;
5. To combine the models with ESBL-E species data and with the WGS isolate typing presented in Chapter 7 to explore carriage at the level of species, *E. coli* sequence cluster and ESBL-containing mobile genetic element.

## 8.3 Methods

### 8.3.1 Developing the models used in this chapter

In the broadest sense when constructing a model, our aim is to estimate the most likely values of the parameters of the model,  $\theta$ , given the data we have,  $x$ . The starting point for estimating likely parameter values, given a choice of model, is usually the *likelihood*: this is the probability of the data, given a set of parameter values. In standard probability notation, this is written as  $P(x|\theta)$ . In fact, this is not the quantity we are interested in; we would like to know  $P(\theta|x)$ : the probability of the parameter values, given the data. Both frequentist and Bayesian modelling approaches provide methods to estimate this quantity, but the starting point for both is the likelihood,  $P(x|\theta)$ , because it is usually much more straightforward to derive an expression for  $P(x|\theta)$  rather than  $P(\theta|x)$ . I will here derive a general likelihood for a two state intermittently observed process; in order to use this likelihood, it is necessary to make some assumptions about the data generating process. I have chosen to use a Markov model, and I will then derive the likelihood for this model, describe how covariates will be incorporated, describe how the model was fit - the process taking us from the likelihood to the most likely parameter values - and finally how goodness of fit was assessed.

### 8.3.2 General form of likelihood

First, I derive a general expression for the likelihood of a two-state intermittently observed process without making any assumptions about the model structure or functional form. Assume we have  $N$  participants with any given participant  $n$  in a state  $S_n(t)$  at time  $t$ : either ESBL-E colonised ( $S_n(t) = 1$ ) or uncolonised ( $S_n(t) = 0$ ). For each participant  $n$  we have a number of measurements of  $S_n(t)$  at a number of time points. The number of measurements varies for each participant, and can be denoted by  $j_n$ , making the time of measurements  $t_{j_n}^n$  for participant  $n$ ; and so for each participant we know the  $j_n$  values  $S_n(t_{j_n}^n)$ .

To arrive at the likelihood for these observations, consider first the simplest situation that we have: the measurements of ESBL status at two time points,  $t_A$  and  $t_B$  for a single participant,  $n$ . The likelihood we wish to calculate, in words, is the probability of the participant being in the second observed state at time  $t_B$ , given they were in the first state at  $t_A$  and given the parameters of the model,  $\theta$ . Or, mathematically:

$$P(S_n(t_B)|S_n(t_A), \theta) \tag{8.1}$$

Assuming all the observations are independent, the probability of all of the states we have

observed for this participant is the product of all the probabilities of the individual states:

$$\prod_{k=2}^{j_n} P(S_n(t_k^n) | S_n(t_{k-1}^n), \theta) \quad (8.2)$$

And the probability of observing the data we have is then simply the product of the probability of all the individual transitions:

$$\prod_{n=1}^N \prod_{k=2}^{j_n} P(S_n(t_k^n) | S_n(t_{k-1}^n), \theta) \quad (8.3)$$

This is the quantity that we wish to calculate: the likelihood for the observed data,  $P(x|\theta)$ . Note that the sum over states for an individual in equation (8.3) starts from 2; if a participant has only one available sample then this does not provide any information about transition probabilities, and must be excluded from the analysis.

### 8.3.3 Markov model likelihood

In order to calculate the likelihood, we need to make some assumptions about the data generating process. In this case, I have chosen to use a Markov model. Markov models are defined by instantaneous transition probabilities, analogous to the hazard of death in a survival model, which is a simple two-state Markov system. Unlike a survival model (where it is not possible to move from the death state to alive), a general Markov model is defined by a transition hazard from each state to each other state in the system. These are traditionally expressed as a Q matrix of instantaneous transition intensities[377,381] (assuming a two-state system):

$$\mathbf{Q}(t) = \begin{pmatrix} q_{00}(t) & q_{01}(t) \\ q_{10}(t) & q_{11}(t) \end{pmatrix} \quad (8.4)$$

Where  $q_{ij}$  represents the instantaneous transition intensity from state  $i$  to state  $j$ . The rows of the Q-matrix must sum to 1 (every participant has to be in one state or another), so if we define the hazard of ESBL-E acquisition to be  $\lambda$  and the hazard of ESBL-E loss to be  $\mu$  (8.1), the Q-matrix becomes, in our case:

$$\mathbf{Q}(t) = \begin{pmatrix} -\lambda(t) & \lambda(t) \\ \mu(t) & -\mu(t) \end{pmatrix} \quad (8.5)$$

However, we are not interested in the Q-matrix *per se* but rather the probability  $p_{ij}$  of starting in state  $i$  at time 0 and being in state  $j$  at time  $t$ ; this can be written in matrix notation as  $\mathbf{P}(t)$  and is related to  $\mathbf{Q}(t)$  by the differential equations:

$$\frac{d\mathbf{P}(t)}{dt} = \mathbf{Q}(t) \cdot \mathbf{P}(t) \quad (8.6)$$

Where  $\mathbf{Q}(t) \cdot \mathbf{P}(t)$  is the matrix product of  $\mathbf{Q}(t)$  and  $\mathbf{P}(t)$ . In order to evaluate  $\mathbf{P}(t)$ , therefore we need to solve this system of differential equations. However, there are limited situations in which these equations have analytic solutions. If the system has time constant or piecewise constant  $\mathbf{Q}$  matrix the matrix exponential is a solution:

$$\mathbf{P}(t) = e^{\mathbf{Q}t} \quad (8.7)$$

However, there is no reason to suspect particularly that the effect of covariates on ESBL-E carriage (e.g. antimicrobials) would be stepwise constant and so a more flexible model is needed. For general time-varying transition intensities, there is no analytic solution to the above equations. However, all is not lost: we can express the likelihood in terms of the differential equations defined by the equations above and solve them numerically in order to calculate the likelihood. The matrix notation above can be simplified, assuming that the system starts in state 1 or 0:

$$\frac{dP_0(t)}{dt} = -\lambda(t)P_0(t) + \mu(t)P_1(t) \quad (8.8)$$

$$\frac{dP_1(t)}{dt} = \lambda(t)P_0(t) - \mu(t)P_1(t) \quad (8.9)$$

Where  $P_i(t)$  is the probability of being in state  $i$  at time  $t$ . Numerical ordinary differential equation (ODE) solvers can quickly solve these equations to calculate, for example,  $P(S_n(t_B)|S_n(t_A), \theta)$  from the simplest example above: the probability that a participant  $n$  at time  $t_B$  is in a given state, given that they were in state  $S_n(t_A)$  at time  $t_A$ , and given the parameters  $\theta$ . This calculation can be completed for all measurements and participants, resulting in the likelihood of the system,  $P(x|\theta)$ .

In order to use this model for inference, two questions must be addressed: first, how to incorporate time-varying covariates; and second, how to practically fit the model. I address each of these questions below.

#### 8.3.4 Incorporating covariates: a proportional hazard model

I have chosen to incorporate covariates using a proportional-hazards model, following both Marshall and Jones[382] and the *msm* package in R[381]. In this model the transmission intensities become:

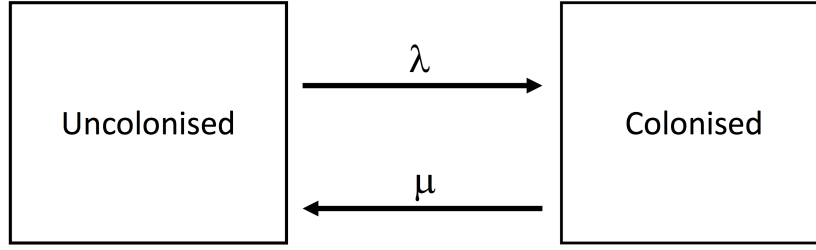


Figure 8.1: Two state ESBL-E model showing instantaneous hazard of ESBL-E acquisition ( $\lambda$ ) or loss ( $\mu$ ).

$$\lambda(t) = \lambda_0 \exp(\beta_0 x_0(t) + \beta_1 x_1(t) + \dots \beta_m x_m(t)) \quad (8.10)$$

$$\mu(t) = \mu_0 \exp(\alpha_0 x_0(t) + \alpha_1 x_1(t) + \dots \alpha_m x_m(t)) \quad (8.11)$$

Where the  $x_k, k = 1, 2 \dots m$  are the  $m$  time-varying covariates in the model and the coefficients  $\alpha_k$  and  $\beta_k$  are the coefficients of these covariates; these have a straightforward interpretation in that the exponential,  $e^{\alpha_k}$  or  $e^{\beta_k}$  can be interpreted as a hazard ratio, as per a simple survival model.

An assumption then needs to be made about the functional form of  $x_m$ . In a stepwise-constant covariate model in which an exposure occurs between  $t_A$  and  $t_B$ ,  $x(t)$  would take the value 1 for all  $t_A \leq t \leq t_B$  and 0 at other times, meaning that the effect of the exposure does not persist once it ceases. Though this may be plausible for some exposures, it seems possible that antimicrobial exposure (for example) might have a longer lasting effect or post-exposure effect; in order to explore this possibility, it is necessary to decide on a flexible, plausible, functional form that such an effect might take. I have decided to use an exponential function, such that:

$$x_k(t) = \begin{cases} 0 & \text{if } t < t_A \\ 1 & \text{if } t_A \leq t \leq t_B \\ \exp \frac{-(t - t_B)}{\gamma_k} & \text{if } t > t_B \end{cases} \quad (8.12)$$

Where the parameter  $\gamma_k$  is a model parameter for each of the covariates, to be estimated from the data, and is related to the half life,  $t_{\frac{1}{2}}^k$  of the decay of the effect of the exposure by:

$$t_{\frac{1}{2}}^k = \gamma_k \ln(2) \approx 0.69 \gamma_k \quad (8.13)$$



This parametrisation has the advantage that the data can fit the size of the parameters  $\gamma_k$ ; if the data are more in keeping with a stepwise effect of the covariates, then a small ( $\ll 1$ )  $\gamma$  would approximate a step function and this could be fit by the model. Alternatively a larger  $\gamma$  would result in the effect of the covariate persisting after exposure, but decaying over time. This allows us to test the hypothesis that antimicrobial exposure (for example) has an effect that persists once exposure finishes, by both the magnitude of the fitted  $\gamma_k$ , and comparing stepwise-constant covariate models to models incorporating the  $\gamma_k$  parameters.

The parameters of the model all have the advantage of having a reasonably intuitive meaning:  $\exp(\alpha)$  and  $\exp(\beta)$  are the hazard ratio for ESBL-E loss and acquisition, respectively; the reciprocals of  $\lambda$  and  $\mu$  are the mean time in days spent in the uncolonised or colonised states, respectively, with covariate values set to 0; and  $\ln(2)\gamma \approx 0.69\gamma$ , as stated above, is the half life of the post-exposure effect.

### 8.3.5 Building and fitting models

The Bayesian probabilistic programming language Stan incorporates an ordinary differential equation solver, and will allow the fitting of the model in a Bayesian framework[320]. In this framework, Bayes' rule allows us to estimate our probability distribution of interest,  $P(\theta|x)$ , called the *posterior* in the Bayesian framework, as long as we provide a *prior*, encoding our prior beliefs about the values of the parameters as a probability distribution for each parameter[316]. Stan then uses the No-U-Turn Sampler (NUTS) implementation of Markov-chain Monte-Carlo (MCMC) sampling[383] to sample from the posterior to provide  $P(\theta|x)$ . It can be shown that, given infinite chain length, MCMC estimates are guaranteed to be unbiased samples from the posterior; when this occurs the chains have said to converged. Unfortunately there is no diagnostic test that guarantees convergence, rather tests that are necessary but not sufficient to ensure convergence: running multiple chains from different starting points with examination of traceplots to show within and between mixing of chains, and the  $\hat{R}$  statistic, which measures mixing of the two halves of an MCMC chain. At convergence,  $\hat{R}$  should be close to 1[316]. In addition, divergences - failure in the NUTS sampler - can be indicative of difficult topography in the posterior at the area where the divergences occur, and suggest that parameter estimates may be biased[384], and are flagged by Stan. All of these tests were used to diagnose convergence.

Two decisions must be made in order to fit the model: covariates must be chosen to include, and priors specified. Models were built sequentially to predict ESBL-E status, starting from the simplest possible, then adding complexity:

- *Model 1:* Composite antibacterial variable (includes all antibacterials) and hospitalisation variable as explanatory variables, both included with stepwise constant effect and no post exposure effect.

- *Model 2:* As per model 1 except antibacterial exposure modelled with decaying post-exposure effect.
- *Model 3:* Hospitalisation, TB therapy and co-trimoxazole exposure all modelled as stepwise constant covariates. All other antibacterials included in a composite variable with decaying post-exposure effect.
- *Model 4:* Hospitalisation, TB therapy and co-trimoxazole exposure all modelled as stepwise constant covariates; ceftriaxone, ciprofloxacin and amoxicillin exposure included in a composite variable with decaying post-exposure effect, with  $\gamma$  allowed to vary for each agent.

Weakly informative priors were used. A normally distributed prior centred at 0 with standard deviation 2 was used for all the  $\alpha$  and  $\beta$  parameters. A parameter value of 2 corresponds to a hazard ratio of 7.4; it would be surprising if any effect is greater than this so this could be argued to be a weakly informative prior. Normally distributed priors centred at 0 with standard deviation 0.2 were used for the  $\mu$  and  $\lambda$  parameters; in a model with no covariates, the inverse of these parameters are the mean times that an individual would remain in the colonised or uncolonised states, respectively, so a value of 0.2 corresponds to a mean state occupancy time of 50 days. A normally distributed prior centred at 0 and with standard deviation 50 days was used for all  $\gamma$  parameters.

The Stan code for the models is given in the appendix to this chapter. Four chains were run in each case, with a warmup of 500 iterations and run for 1000 iterations in total. Convergence was assessed using the diagnostics described above. Stan v2.19 was used to sample from the posterior, accessed via Rstan v2.19.2, and run on the Wellcome Sanger Institute computing cluster under Linux Red Hat v7.6, running R v3.5.3, and gcc v6.3.0 C++ compiler. Four cores (one per MCMC chain) and 3GB of memory per model fit were used. Posterior samples were brought to my local machine (MacBook Pro running mac OS Mojave 10.14.5) and further analyses undertaken with R3.6.0.

### 8.3.6 Assessing goodness of fit

Model goodness of fit was assessed in two ways; first, by graphical posterior predictive checks: comparing predicted total number of ESBL-E positive samples to the actual number across the three arms. This was done by using the posterior parameter estimates for each MCMC draw (after discarding warmup samples) to generate a predicted probability of the ESBL-E positive state for each data point, then sampling from a Bernoulli distribution to convert to predicted state occupancy. Each data point therefore had 2000 predictions for state occupancy, one for each posterior draw. These were plotted as kernel density plots against actual state occupancy, stratified by arm, to visualise the goodness of fit of the model, and to compare between models.

Second, models were compared using leave-one-out cross validation, as implemented in the *loo* v2.1.0 package in R[322]. This estimates the out-of sample predictive ability of the model by estimating a quantity called the expected log predictive density (*ELPD*) essentially the log of the likelihood for a new, unseen dataset conditional on the current data. This quantity is estimated using leave-one-out cross validation to produce and estimate of the *ELPD*, hereafter referred to as *ELPD<sub>loo</sub>*. The standard error of *ELPD<sub>loo</sub>* for a model is also calculated and so two models can be compared by comparing the *ELPD<sub>loo</sub>* difference and standard error; if the difference is greater than twice the standard error (i.e. a 95% confidence interval, assuming normality) we can be confident that one model would be expected to have greater out-of-sample predictive ability than the other[322]. Because this technique estimates out-of-sample predictive ability it naturally incorporates a penalty for including multiple parameters and hence overfitting, as an overfit model would be expected to have worse out of sample predictive ability and hence lower *ELPD<sub>loo</sub>*.

### 8.3.7 Exploring differences in carriage dynamics by bacterial species and *E. coli* genotype

The models fit as described above predict whether a participant will be colonised with any ESBL producing organism at a given time point, but this classification obscures a lot of complexity. A participant can be colonised with different ESBL- producing species (largely *Escherichia coli* or *Klebsiella pneumoniae*), and different clones of those species containing different ESBL genes on different mobile genetic elements (MGEs). It may be that there is heterogeneity in carriage dynamics across these different levels of the system. To address this hypothesis, the best fitting model identified from the four described above was refit but the “colonised” state modified to either consider the species level or to use the whole genome sequence data presented in Chapters 6 and 7. The analysis in Chapter 7 suggests that the element most conserved within participants is the bacterial clone-ESBL contig combination, where the bacterial clone clusters were defined with the hierarchical BAPS algorithm and the ESBL-contig clusters defined with the cd-hit algorithm, as described in Chapter 7. The hierBAPS cluster-contig cluster pairs are coded as follows in this chapter: a.ESBLgene.b where a is the ID number of the level 2 hierBAPS cluster, and ESBLgene.b is the number of the contig cluster for a given ESBL gene, and for the rest of the chapter for brevity each unique hierBAPS cluster-contig cluster will be referred to as an *E. coli* genotype. All *E. coli* genotypes which were identified in more than 15 samples - 6 in total - were included and so the models were refit defining the colonised state as the presence of, respectively, ESBL *E.coli* , *K. pneumoniae* or one of the six included hierBAPS cluster-contig cluster pairs. The parameters for these models were compared with each other and with the original ESBL model.

### 8.3.8 Simulations from the posterior

Finally, in order to better understand the relative role of antimicrobial exposure and hospitalisation in driving ESBL-E carriage, I conducted simulations with these exposures set at varying levels. The probability of ESBL colonisation as a function of time was calculated by solving the equations (8.8) and (8.9) using the R package *deSolve* v1.2.4[385], for each of the 2000 posterior parameter estimates from the posterior and assuming a 50% initial probability of ESBL colonisation. This yielded a distribution of carriage probability at each time point which was summarised using the median and 95% confidence intervals and plotted against time for varying covariate values: days of hospitalisation was varied from zero to twenty in steps of five, as was antimicrobial exposure, and each simulation repeated both with and without CPT.

## 8.4 Results

### 8.4.1 The effect of antibacterials and hospitalisation on ESBL-E carriage

First, I fit the four models with three specific aims:

1. To identify the model that provides the best trade off between predictive ability and the computational cost to fit;
2. To explore the relative effects of hospitalisation versus antimicrobial exposure on ESBL-E carriage by assessing the posterior parameter values of these models;
3. To assess support in the data for a post-antibacterial effect on ESBL-E carriage that persists once antibacterial therapy is stopped.

For these models, the colonised state was defined as at least one ESBL producing organism of any species identified in a sample, and uncolonised as no ESBL producer identified. After excluding participants with only one sample, there were 993 pairs of samples in 363 participants remaining that contributed data to the analysis. All four models converged within the 1000 iterations;  $\hat{R}$  was less than 1.1 for all parameters and all traceplots showed good mixing of chains. There were no divergences of the NUTS MCMC sampler in any of the models. There was a computational cost to increasing the number of parameters, as would be expected from the increase in dimensionality of the posterior: model one took 3.5 hours to fit, model two 13.7 hours, model three 17.1 hours and model four 33.4 hours.

The parameter estimates for the models are shown in Figure 8.2. There were significant correlations between some posterior parameters (see Figure 8.6 in the chapter appendix for pairwise plots for model 2 as an example): particularly  $\lambda$  and  $\mu$ , and the  $\alpha$  and  $\beta$  parameters. This is not necessarily problematic in that it is not necessarily a source of

bias, but can make it difficult for some MCMC algorithms (e.g. Metropolis-Hastings) to adequately sample from the posterior[316]. Nevertheless, the diagnostics suggest that the Stan NUTS sampler had no problems.

The effect of hospitalisation is consistent across all models; in most models, the 95% credible intervals for both  $\alpha$  and  $\beta$  for hospitalisation do not cross zero and are positive, suggesting that the hazard ratio of hospitalisation on both the rate of acquisition and loss of ESBL-E is very likely to be greater than one, and the effect of hospitalisation is to increase both the rate of acquisition and loss of ESBL-E. The estimated effect sizes are consistent across the models though, as expected, uncertainty in the estimate increases as more parameters are added to the model.

The effect of antibacterial exposure is also reasonably consistent across the models; the parameter  $\alpha$  is negative in all cases, and often the 95% credible intervals do not cross zero, suggesting that the hazard ratio of antimicrobial exposure is likely to be less than zero. The effect sizes are similar in all cases, for all agents (including TB therapy), whether antibacterial exposure is considered as an aggregate variable or as individual agents; though in the extreme case where agents are all considered individually (Model 4, Figure 8.2D) the uncertainty in the estimates makes it difficult to draw any firm conclusions. This suggests that all the considered antibacterial agents act, with broadly similar effect size, to prolong ESBL-E carriage by reducing the rate of loss. No  $\beta$  parameter (the log hazard ratio of ESBL-E acquisition) has 95% credible intervals that do not cross zero, consistent with antibacterial exposure having no or limited effect on ESBL-E acquisition.

The relative predictive ability of the four models were assessed in two ways: first, the predicted proportion of ESBL-E positive samples were plotted by sampling from the posterior (Figure 8.3); second, the pairwise  $ELPD_{loo}$  differences (and standard errors in the differences) between all models calculated (Table 8.1). All models predicted ESBL-E carriage reasonably poorly for arm two and three participants, but better for arm one (Figure 8.3). The addition of a post-antibiotic effect improved model fit (seen by comparing model 1 to model 2 graphically in Figure 8.3 and by an  $ELPD_{loo}$  difference of more than two standard errors in Table 8.1) but models two, three and four, had similar fit despite the increase in number of parameters from seven in model two to seventeen parameters in model four. Model two therefore provides a good balance between computational tractability, interpretation and predictive ability; the parameter estimates for this model, expressed as hazard ratios for  $\alpha$  and  $\beta$ , the mean time in state for  $\lambda$  and  $\mu$  and half life of post-antibacterial effect for  $\gamma$  are shown in Table 8.2. The significant  $ELPD_{loo}$  difference between model 1 and 2 provides strong support for a post-antibiotic effect of antimicrobial exposure.

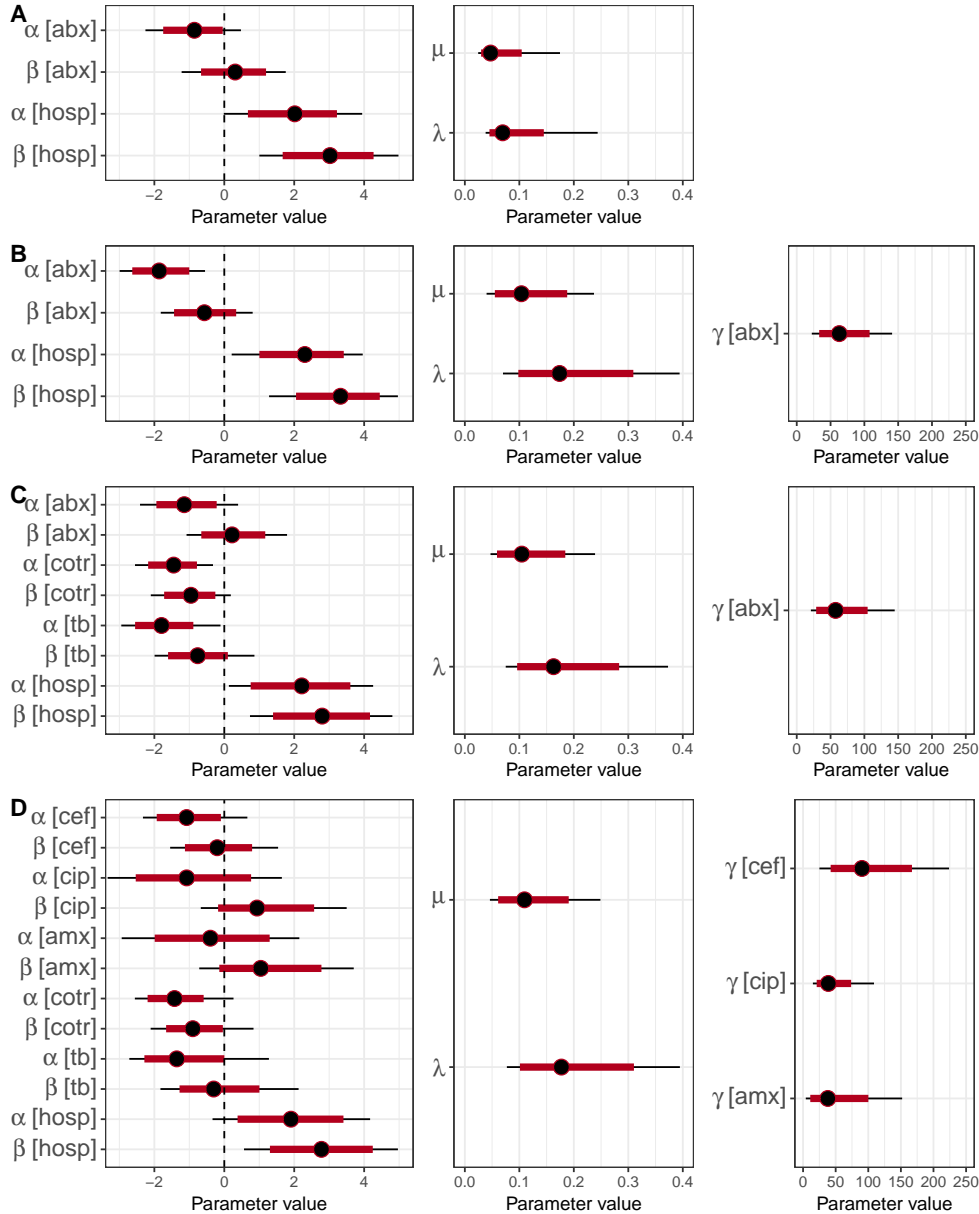


Figure 8.2: Parameter estimates from increasingly complex Markov models to predict ESBL carriage. Black lines are 95% and red lines 80% credible intervals. A: Model 1 includes stepwise constant covariates only, antimicrobial exposure (abx) and hospitalisation (hosp).  $\lambda$  is the baseline hazard and  $\beta$  the log hazard ratio of ESBL-E acquisition,  $\mu$  the baseline hazard and  $\alpha$  the log hazard ratio of ESBL-E loss. B: Model 2 adds a post-exposure effect of antimicrobial exposure, parameterised by  $\gamma$  as described in the text. C: Model 3 adds stepwise constant covariates for TB therapy (tb) and cotrimoxazole (cotr) with all other antimicrobial exposure captured in the abx variable, which has a post exposure effect as before. D: Model 4 separates the effect of antimicrobial exposure into the component agents, with post exposure effects for all except cotrimoxazole and TB therapy. In most models 95% credible intervals of  $\alpha$ [hosp] and  $\beta$ [hosp] do not cross zero and are positive, suggesting that hospitalisation acts to both increase rate of ESBL-E acquisition and loss; for antimicrobial exposure, on the other hand, only the 95% for antimicrobial  $\alpha$  values consistently do not cross zero, and are negative, suggesting that the effect of antimicrobial exposure is to reduce the rate of ESBL-E loss. It is also clear that adding parameters to the model increases the uncertainty in the estimates (e.g. compare model 2, B, to model 4, D).

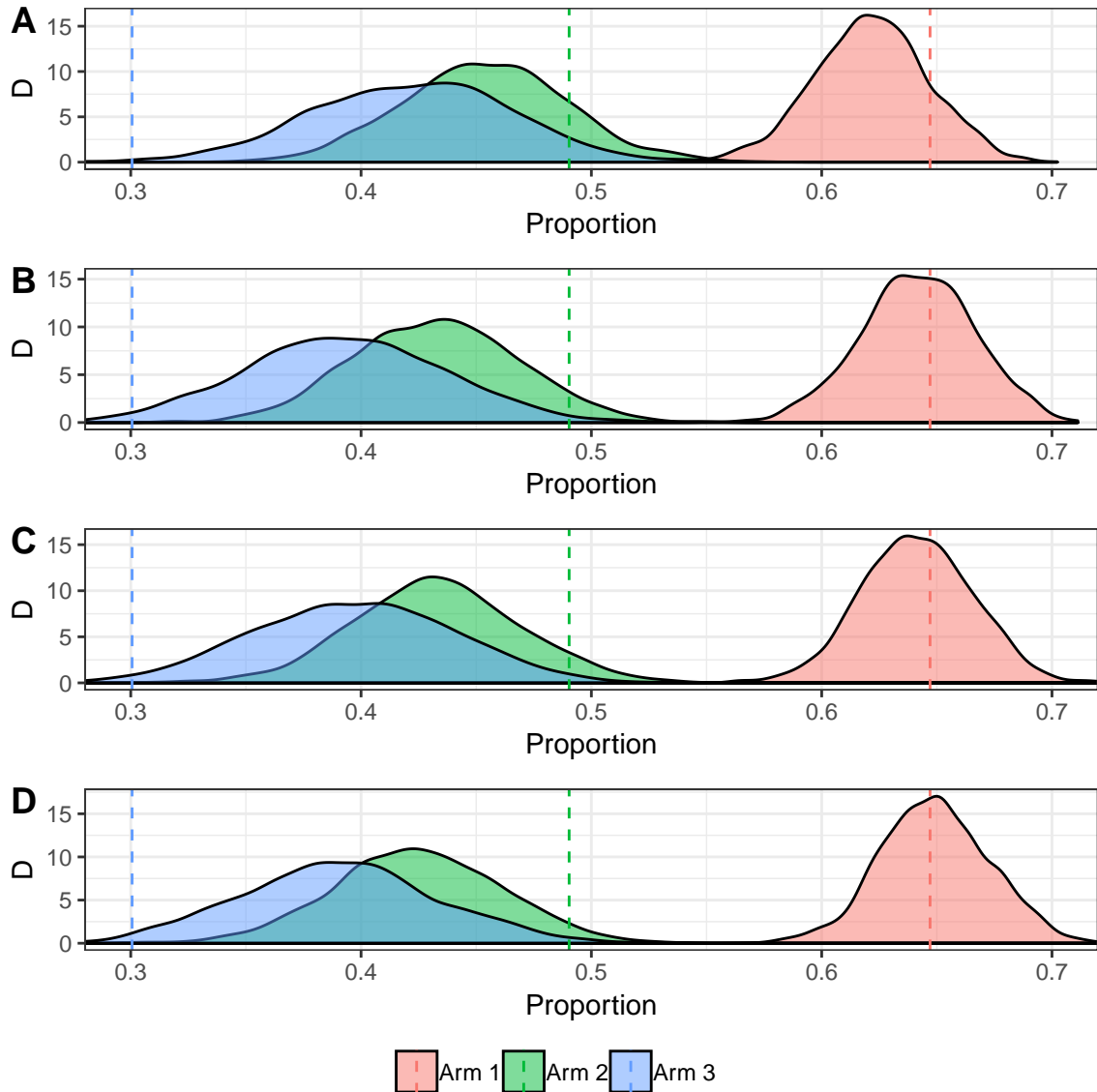


Figure 8.3: Posterior predictive checks: kernel density estimate,  $D$ , of predicted proportion of ESBL-E positive samples, stratified by arm for Model 1 (A), Model 2 (B), Model 3 (C) and Model 4 (D), generated by sampling from a Bernoulli distribution using the predicted probability for each sample ( $n=993$ ) for each draw from the posterior, excluding warmup draws ( $n = 2000$ ). True proportion of ESBL-E positive samples are shown for each arm by dotted vertical line. In all cases, predictions are poor for arm 2 and 3 samples, but the addition of a post-antibacterial effect improves fit, especially in arm 1 participants: compare Model 1 (A) with stepwise constant covariates to Model 2 (B) with post-antibacterial effect. Models 2-3 (B-D) have similar predictions despite more parameters.

Table 8.1: Estimates (and standard error) of pairwise expected log predictive density (ELPD) differences for all models

	Model 1	Model 2	Model 3	Model 4
Model 1	0.0 (0.0)	10.5 (4.2)	10.0 (6.4)	15.0 (7.0)
Model 2	-	0.0 (0.0)	-0.5 (5.2)	4.4 (6.0)
Model 3	-	-	0.0 (0.0)	4.9 (3.7)
Model 4	-	-	-	0.0 (0.0)

*Note:*

Cells in table compare row model to column model. A positive number favours the model in the column. The standard error of the ELPD difference is given in brackets; if twice the standard error is less than the estimated ELPD difference then we can be confident that the column model has better out-of-sample predictive fit than the row model. All models have better fit than model 1 but models 2-4 all have similar fit.

#### 8.4.2 Exploring bacterial species and genotype differences in carriage dynamics

Next, I explored the differences in carriage dynamics between ESBL-E species and *E. coli* genotype, by refitting model 2 but considering the colonised/uncolonised states to be, in turn, presence or absence of *E. coli*, *K. pneumoniae* or any of the top six most prevalent *E. coli* genotypes (as defined by the combination of ESBL containing contig cluster and *E. coli* hierBAPS cluster [Chapter 7]), and refitting the model for each one. All 993 within-participant sample-pair comparisons were used to fit the *E. coli* and *K. pneumoniae* models, but because sample collection continued after the sequenced *E. coli* included here were shipped, all samples collected after this time were excluded from the genotype models. 585 samples from 251 participants were therefore included in the genotype models.

The parameter estimates for these eight models (alongside the original ESBL-E presence/absence model) are shown in Figure 8.4. In general, there was more uncertainty in the parameter estimates for the new models, as might be expected as there are fewer carriage events, and fewer samples in the case of the genotype models. The only significant parameter difference between the models was in the  $\lambda$  parameter, the baseline hazard of state acquisition. The magnitude of the difference was large; for example the median (95% CI)  $\lambda_{ESBL}$  estimate of 0.10 (0.07-0.15) is almost three orders of magnitude larger than the estimate of  $\lambda_{6.CTXM.27}$ , 0.002 (0.001- 0.003). These values would correspond to a mean (95% CI) time in the uncolonised state of 10 (6-14) days for the ESBL model versus 500 (333-1000) days for the genotype model, assuming all other covariates were zero. The hazard rate of state loss,  $\mu$  was similar, however, meaning that the time in the colonised state is similar for the ESBL model, and for all the *E. coli* genotype models.



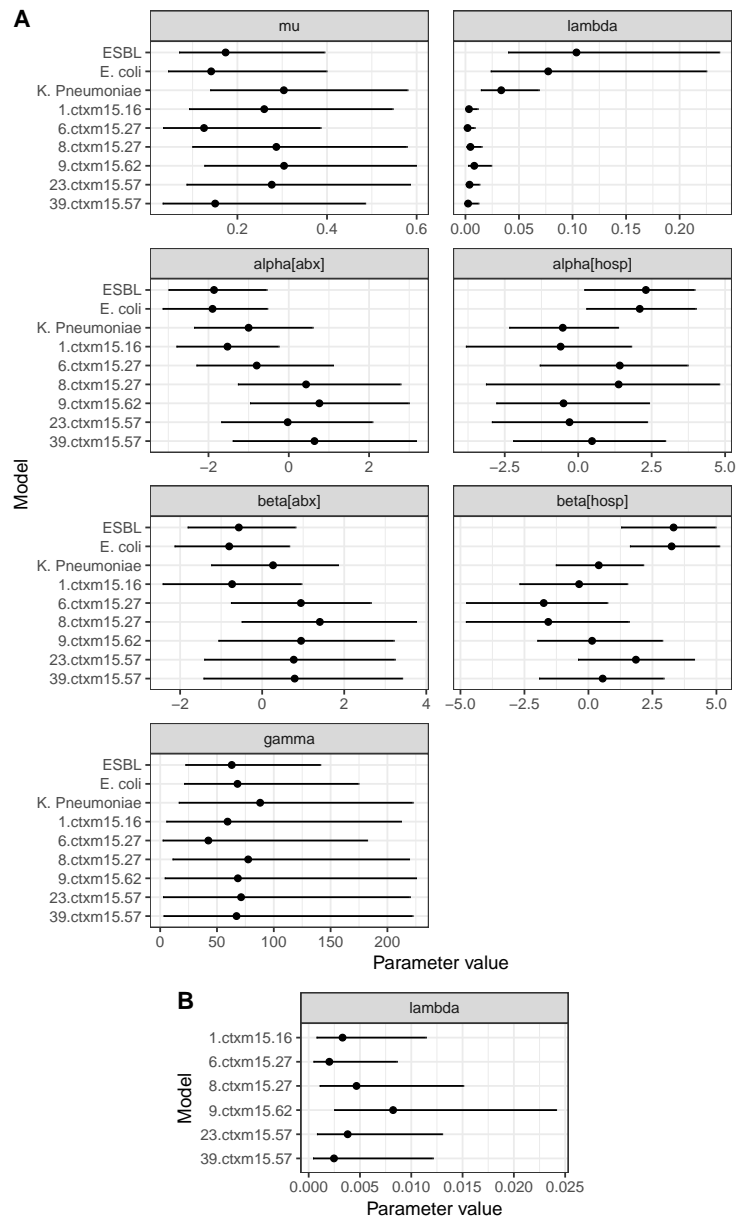


Figure 8.4: Parameter estimates from two state models predicting species and *E. coli* genotype carriage, compared to original model, which predicted carriage of any ESBL-*E*. A: All parameters, showing that the only significant difference between the models is the parameter  $\lambda$  (the hazard of acquisition), with an order of magnitude difference between the hazard of ESBL acquisition versus the acquisition of a particular genotype. B:  $\lambda$  parameter only for genotype models, showing that the estimates are similar for each genotype.

Table 8.2: Parameter estimates (and 95% confidence intervals) from model 2

Variable	Value
<b>Effect of Antibacterials</b>	
Hazard ratio for ESBL-E loss	0.16 (0.05-0.58)
Hazard ratio for ESBL-E acquisition	0.57 (0.16-2.25)
<b>Effect of Hospitalisation</b>	
Hazard ratio for ESBL-E loss	10.01 (1.24-52.34)
Hazard ratio for ESBL-E acquisition	27.82 (3.60-143.18)
<b>Post Antibacterial Effect</b>	
Half life (days)	43.67 (15.42-97.66)
<b>Mean time in state</b>	
Uncolonised (days)	9.65 (4.22-25.07)
Colonised (days)	5.76 (2.54-14.30)

*Note:*

Hazard ratios are the exponential of the parameters  $\alpha$  and  $\beta$  in the model; half life is equal to  $\log(2)$  multiplied by  $\gamma$ ; mean time in state assumes all other covariates are equal to zero and is then the reciprocal of  $\lambda$  or  $\mu$ .

### 8.4.3 Simulation of different antibacterial and hospitalisation scenarios

Finally, to better understand the relative roles of antimicrobial exposure and hospitalisation in driving ESBL-E carriage, I simulated the probability of ESBL-E colonisation as antibacterial and hospital exposure changed from 0 to 20 days, assuming a 50% baseline probability of ESBL-E colonisation (Figure 8.5) and both with and without cotrimoxazole preventative therapy. Model 2 was used for these simulations. Hospitalisation seems to rapidly increase carriage probability and antimicrobial exposure produces a slower rise. Most striking, however, is that the effect of both exposures simultaneously causes a rapid increase in ESBL-E colonisation probability as well as prolonged decay to baseline probability: by the end of the 100 day simulation period in those simulations with both hospital and antibacterial exposure, most probabilities have not yet returned to baseline levels. Shorter course lengths of antibacterials seemed to have similar effects to longer courses. In the model used for the simulations (model two), the effect of all antibacterials (including CPT) is equal and so CPT seems to be the primary driver of an increased long-term carriage probability. TB therapy is also included in the composite “antibacterial” variable, so these conclusions would be equally valid for TB therapy.

## 8.5 Discussion

In this chapter, I have extended the continuous-time Markov models available in the *msm* package in R to incorporate true time-varying covariates (rather than stepwise constant). I have fitted them to the data presented in Chapter 5 using a Bayesian framework and a

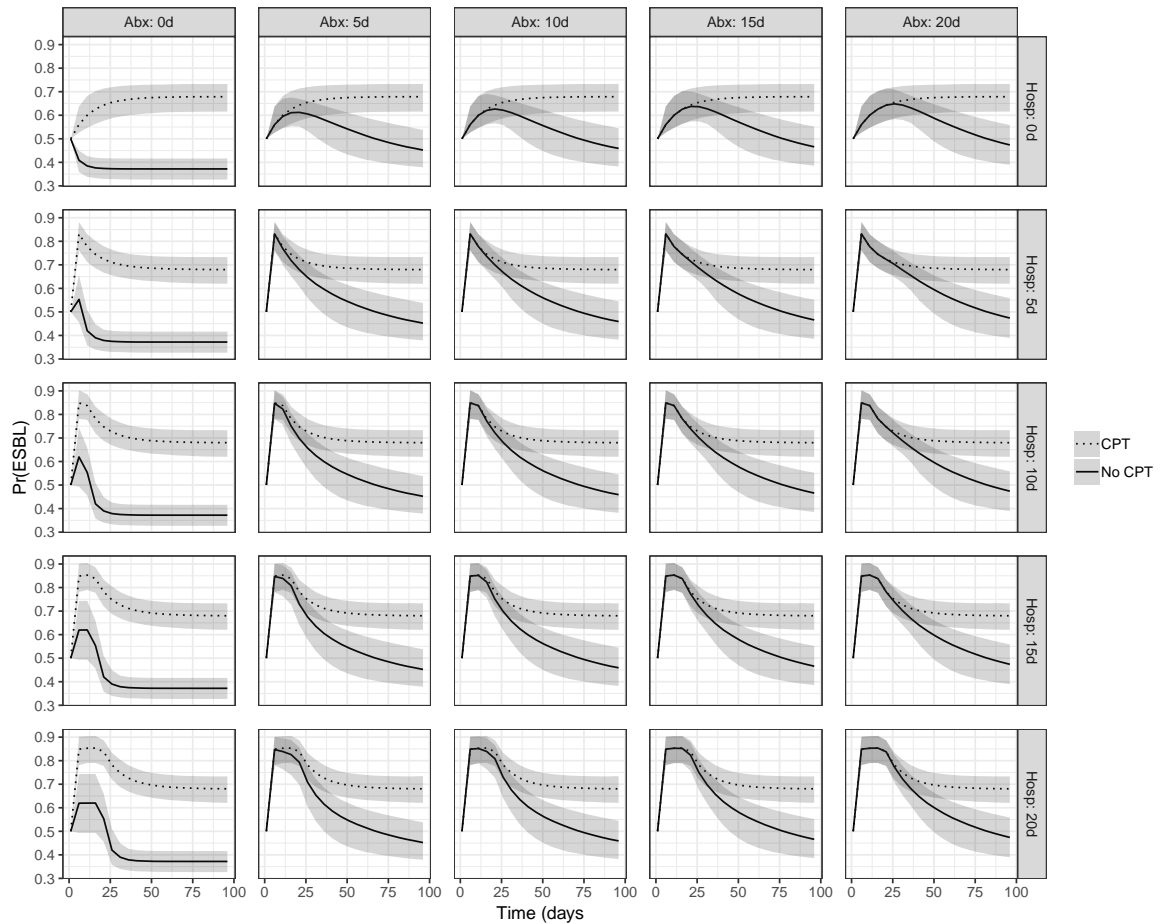


Figure 8.5: Simulations of different antibacterial and hospitalisation scenarios. CPT = Cotrimoxazole preventative therapy. Plots show estimated probability of being in the ESBL+ state for given covariate values as a function of time, assuming a baseline 50% probability of ESBL-E colonisation. Antimicrobial exposure ranges across columns from 0 to 20 days, and hospitalisation across rows from 0 to 20 days. Hospitalisation causes a more rapid increase in probability than antimicrobial exposure, whereas antimicrobial exposure in the form of CPT is the primary determinant of increased long-term carriage probability. However, there is a synergistic effect of hospitalisation and antimicrobial exposure that results in rapid rise in colonisation probability and prolonged decay.

differential equation solver in the probabilistic programming language Stan. From these fitted models, it is possible to draw several conclusions.

First, the class of models that I present are feasible to fit in a reasonable amount of time with modest computational requirements, and are very flexible. The models were largely fit overnight on the WSI cluster with four cores and 3GB RAM. These are not particularly onerous requirements, and the times to fit would be expected to be similar on a desktop machine. The parametrisation of the model is extremely flexible; I chose an exponential form of a post-antibacterial effect but any functional form could be used, simply by replacing the function that generates the covariate values,  $x(t)$  in Stan model. If a function can be written down, it can be fitted in this framework with minimal effort. This provides, for example, the opportunity to explore *in silico* different hypotheses as to the ways in which antimicrobial exposure drives ESBL-E carriage, by exploring the functional form of the antimicrobial exposure covariate that best fits the data.

Second, the values of the parameter estimates and the simulations from the ESBL models allow an insight into the drivers of ESBL-E colonisation in Malawian adults, and suggests areas to target for interventions. Hospitalisation acts to increase both ESBL-E acquisition and loss, the net result of which is a rapid increase in the probability of ESBL-E colonisation following admission. Antimicrobial exposure acts to reduce the rate of ESBL-E loss and thus prolong carriage, but it simultaneous hospitalisation and antimicrobial exposure have a synergistic effect to produce the observed rapid increase in ESBL-E carriage prevalence seen in antibacterial exposed inpatients. The crude unadjusted analysis suggested that antimicrobial exposure alone was driving ESBL carriage; in fact, these models show that both hospitalisation and antimicrobial exposure act in synergy.

Whether these associations represent causal relationships and the mechanisms of any relationships are not addressed by the models, but causal associations are certainly biologically plausible. The hospital environment at QECH is such that cleaning is difficult, hand washing facilities for staff, participants and guardians are lacking, wards are crowded with participants close together, and one toilet is shared between around 60 patients, all of which potentially facilitate the acquisition of ESBL-E. The level of granularity of the models is such that the relative contributions of person-to-person transmission or transmission from the hospital environment or toilet facilities is not clear, and as it stands hospitalisation is used as an aggregate proxy for all of these putative transmission routes.

The genomics data (Chapter 6) perhaps provides additional insight. There is no one hospital clone and in terms of *E. coli* diversity at least, the hospital is largely an extension of the community. Given the number of adult admissions to QECH per year - around 15,000[386], this is perhaps not surprising: if each adult admission is cared for by two guardians then ~5% of the population of Blantyre - 800,000 at the 2018 census - is passing through QECH yearly. Given this, there is likely a frequent influx of ESBL-E clones from

the community, which could, given the genomic data, suggest simple patient-to-patient transmission is the main route of ESBL-E acquisition; however, the hospital association of *E. coli* ST410 implies a hospital reservoir of some sort. Further studies are needed to clearly understand transmission dynamics of AMR-bacteria within the hospital setting in Malawi, which could inform infection prevention and control (IPC) interventions. Evidence based IPC interventions which can be deployed in low-resource settings are urgently needed and, given the rates of hospital exposure in Blantyre, could potentially have a significant impact on the prevalence of ESBL-E there.

Strategies to mitigate against the effect of antimicrobials on ESBL-E carriage are also needed. The data presented here support a post-exposure effect of antibacterials on prolonging ESBL-E carriage duration, such that short courses of antimicrobials seem to have a similar effect to longer courses in hospitalised participants. This finding may be contingent on the parametrisation of the post-antibiotic effect, and requires further exploration, but could have significant implications for antimicrobial stewardship. In this model framework, two days of antibacterial therapy to ten inpatients would result in more participant-days carriage of ESBL-E than twenty days of antibacterial therapy to one patient - as seen from the simulations presented in Figure 8.5 - despite the same number of defined daily doses being used in total. This would suggest that antimicrobial stewardship interventions to avoid unnecessary antibacterials altogether would be more effective than those limiting antibacterial course lengths by e.g. review of blood culture results at 48 hours. The post-antibacterial effect has a lengthy half life of 44 days (95% credible interval 15-98 days), much longer than the time by which most antimicrobials will have been excreted from the body. Such a prolonged effect is biologically plausible, however even short courses of antimicrobials are known to profoundly alter the composition of the gut microbiota[387,388], which could certainly alter ESBL-E carriage dynamics[389]. Further studies of the role of the microbiota in colonisation resistance to ESBL-E could shed light on the mechanisms of the post-antibacterial effect I demonstrate here, and pave the way for microbiota-modulating therapies to mitigate against it.

The role of CPT in driving long-term ESBL-E carriage is likely significant, and it appears to be a major determinant of long-term ESBL-E colonisation. Again, this is perhaps not surprising given that cotrimoxazole exposure dwarfs exposure to all other antimicrobials combined in the cohort, and that cotrimoxazole resistance genes are near-ubiquitous in the sequenced ESBL *E. coli* isolates. It is not possible to say from the short read sequence data but cotrimoxazole resistance determinants could be collocated with ESBL resistance determinants, which would explain an apparent selection pressure for ESBL-E carriage from CPT. CPT has been shown to have significant mortality benefits in people living with HIV[390], and lifelong CPT is mandated by WHO guidelines for all people living with HIV in areas with high malaria or bacterial infection prevalence, including Malawi[330,390]. Given an estimated adult Malawian HIV prevalence of 9.6%[284], CPT is likely therefore a

major driver of ESBL-E carriage in Malawi. The risk of driving AMR with CPT needs to be balanced against its benefits, and may be possible that in the era of high ART coverage, reducing malaria incidence and growing Gram-negative resistance that these risks begin to outweigh the benefits. The exact mechanism by which CPT confers a reduced mortality risk - whether it acts primarily to prevent opportunistic infections, bacterial infections or malaria - remains controversial. A recent RCT in Uganda carried out in 2012[391] showed that a strategy of stopping CPT once the CD4 cell count is persistently above 250 cells  $\mu L^{-1}$  is associated with more CPT-preventable infections, including malaria and pneumonia, but no difference in mortality (1.7% vs 1.8% over 12 months). Results are awaited of the TSCQ trial (ClinicalTrials.gov identifier NCT01650558), which has assessed the effect of mortality of CPT versus chloroquine malaria prophylaxis in Malawian HIV-infected adults, based on the hypothesis that in malaria-endemic areas the mortality benefit of CPT is primarily driven by its antimalarial properties. Given the findings here, a chloroquine based prophylaxis strategy could significantly impact ESBL-E carriage prevalence (and hence, possibly, infections) in Malawi and would be very attractive from this perspective if non inferior to CPT in mortality endpoints.

Finally, using WGS as a high resolution typing tool allows very granular insight into ESBL-E carriage at the genotype level. Within the limitations of reasonably uncertain parameter estimates due to small numbers, all parameters for genotype carriage models were the same as the general ESBL carriage model, with the exception of  $\lambda$ . This indicates that the rate of acquisition of a given *E. coli* genotype is two to three orders of magnitude lower than the overall rate of ESBL acquisition, in turn suggesting that apparent continual ESBL-E carriage in fact represents a much more frequent apparent acquisition of different ESBL-E genotypes. This could represent true acquisition or some other dynamic shift in the relative abundance of ESBL producing *E. coli* in the microbiota over time. This analysis is, however, hampered by the fact that only one colony pick at each time point was sequenced and hence the true distribution of *E. coli* genotypes at a given time point is unknown.

### 8.5.1 Limitations

There are several limitations to the analysis presented here. First, despite a reasonable number of data points, the parameter estimates from these models have moderate uncertainty. Some of this may be consequent on the model structure: with strongly correlated parameters, the data may be consistent with a wide range of paired parameter values. Even those parameter values whose 95% credible intervals cross zero (e.g. the hazard ratio of antibacterial exposure on ESBL-E acquisition in model 2) largely incorporate a clinically meaningful effect size, and so care must be taken not to interpret a lack of certainty of a significant effect as a lack of effect. Uncertainty in parameter estimates increases as more parameters are added, meaning that understanding the relative effects of different

antibacterial agents on ESBL-E carriage is not possible, and in most models antibacterials are considered as an aggregate variable. *A priori*, different antibacterial agents would not be expected to have the same effect on ESBL-E carriage dynamics, but here they are considered together. There is some support for this strategy from the fact that the estimated effect sizes of ceftriaxone, ciprofloxacin and amoxicillin (the most commonly administered antibacterials apart from CPT and TB therapy) are similar when considered individually and in aggregate, but uncertainty in these estimates warrants caution. The apparent effect of TB therapy is particularly surprising, given that the first-line combination of rifampicin, pyrazinamide, ethambutol and isoniazid would be expected to have a limited selection pressure for ESBL-E, and warrants further study. This could represent Gram-negative activity of rifampicin or confounding by CPT, given that the majority of participants receiving TB therapy also receive CPT.

In addition, despite fitting well to participants from arm 1 of the study (those with sepsis), the models fit poorly to arm 2 (antimicrobial unexposed participants) and arm 3 (community members). The reasons for this are not clear, but it strongly suggests that there are covariates that are not included in the model that differentiate the arms of the study in some way. If these covariates are also associated with the exposures of interest (hospitalisation and antibacterial exposure) then this is a potential source of bias from confounding.

These models assume perfect test characteristics, such that the measurement of ESBL-E status (or species or genotype, depending on the model) perfectly represents the true state. This is unlikely to be the case in practice, and there is also likely to be differential test characteristics between the different stool testing methods (stool or rectal swab culture) used. This may have introduced bias to the parameter estimates and simulations. Expanding the model to incorporate imperfect tests - a hidden Markov model - could address this limitation, as well as provide estimates of test sensitivity and specificity. Conceptually this is straightforward; the underlying “true” state is modelled, the likelihood for a given participant (Equation (8.3)) becomes the sum over all possible underlying paths through the system and parameters are added for the sensitivity and specificity of the tests used. This will, however, increase computational costs: if a participant has ESBL-E status measured at  $n$  time points then calculating the likelihood required summing over all  $2^n$  possible combinations of states, rather than just one as in the models presented here.

Generalising the model to allow states to be hidden or censored would also address a serious limitation of the genotype models. In these models, the absence of a particular genotype from a sample is interpreted as true absence, but the true situation is more complex. If no ESBL at all is cultured then we can be confident that a given genotype is absent, within the confines of the test sensitivity. However, if *E. coli* were cultured at any time point, then only one colony pick was taken forward for sequencing, meaning that it is possible that any

number of other genotypes were present in the sample but not picked and sequenced and therefore identified. Data on within-participant gut mucosal ESBL-E diversity are sparse, but those data that are available suggest that it may be considerable[374], and so these models should be considered as merely exploratory. Expanding the model to allow states to be censored (i.e. for the true underlying state to remain unknown for a given measurement) is equivalent to the changes that would be necessary to incorporate hidden Markov models, and would address these problems.

## 8.6 Conclusion and further work

In conclusion, I have developed and fit time-inhomogeneous Markov models to the clinical longitudinal ESBL-E carriage data. The models are computationally tractable, extremely flexible, and provide insight into the drivers of ESBL-E carriage in Blantyre. Though both hospitalisation and antibacterial exposure significantly affect the probability of ESBL-E carriage independently, they appear to act synergistically together to drive colonisation. Antibacterial exposure seems to have an effect that persists long after most antibacterials would be expected to be excreted from the body; the models provide no data on the mechanism of this but one hypothesis would be that it is mediated by changed in the microbiota. Short courses of antibacterials seem to produce a similar effect to longer courses, which may have implications for antibacterial stewardship interventions. Co-trimoxazole preventative therapy may be one of the major drivers of long-term ESBL-E carriage in Malawian adults and this should be considered in developing international and national guidelines on its use.

These conclusions suggest a direction for future work. The models must be expanded to incorporate censored states to allow the fitting of hidden Markov models and to account for the single colony pick sampling method which was used. This, in conjunction with whole genome sequencing of the remainder of the isolates from the study will allow unbiased models to be fitted to understand carriage at the level of the genotype. Finally, shotgun metagenomic sequencing of stored extracted stool DNA from the participants in this study will a) define the total diversity of ESBL genes within each sample and b) will allow an analysis to identify microbiota associations of ESBL-E colonisation and the effect of antibacterial exposure. This will allow testing of the hypothesis that the post-antibacterial effect I have identified is mediated via the microbiota.



## 8.7 Appendix

The Stan code for the fitted models is below; the stepwise-constant covariate model is presented first, and all other models were fitted with the second model. Pairwise posterior parameter estimates for model two (to demonstrate strong parameter correlations) are also shown below; see text for details.

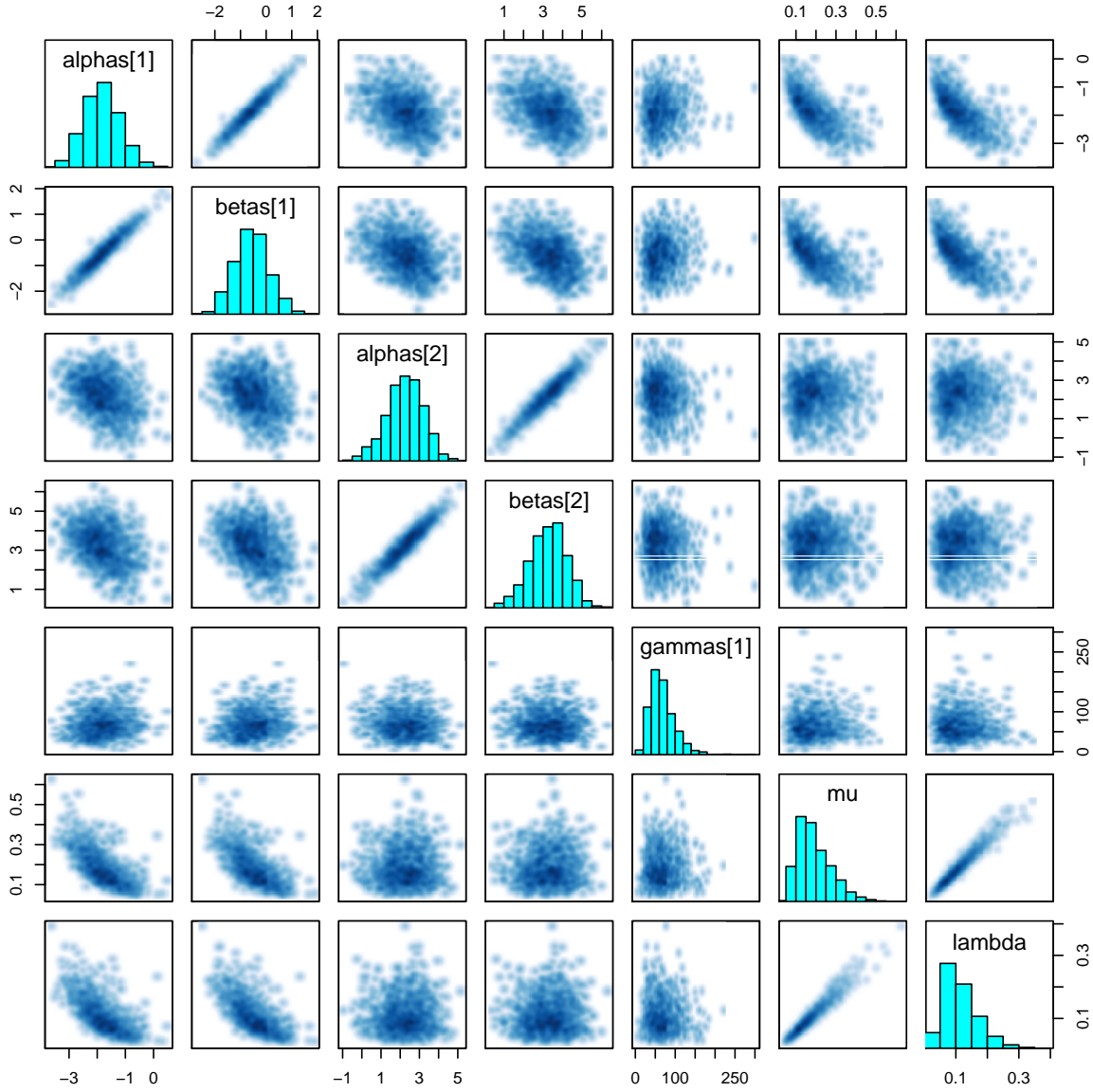


Figure 8.6: ESBL-E carriage model pairwise posterior parameter estimates, showing correlations between alpha and beta for a given covariate, and lambda and mu. These results are for model 2 to predict ESBL-E probability - see text for details. Alpha[1] and beta[1] are the coefficients for the composite antibacterial exposure variable, alpha[2] and beta[2] for hospitalisation, lambda the rate of ESBL-E loss, mu the rate of ESBL-E acquisition, and gamma the scaled (by  $\log(2)$ ) half life of the post-antibacterial effect.

```

// Stan final model incorporating varying numbers of covariates
// Optional gamma decay
// Uses rk45 ODE solver
// Joe Lewis July 2019

// to call this model from Rstan, pass it the following data

// N: integer = number of rows of data, each row consisting of two ESBL
// observations for one patient

// n_covs: integer vector of length 3 = [number of
// nontimevarying covariates,
// number of stepwise constant covariates,
// number of exp decay covariates ]

// covs_type: integer vector of length(number of covariates) =
// each position encodes the type of variable
// in the order they are presented in covs_mat:
// 3 = time varying with exponential decay of effect
// 2 = time varying with piecewise constant
// 1 = nontimevarying
// All the exp decay variables must always go first

// cov_mat: real matrix of start and stop times of
// covariates 3*(with number of covariate) cols
// Each covariate needs three columns, in this order
// start_time: time that covariate started
// stop time: time that covariate stopped.
// If there is no covariate exposure in this row, code as -999
// prev_stop time: if covariate has exp. decay, this is
// the previous stop time (before current row e.g. -10)
// If no previous exposure, code as 999
// If non time varying exposure, code as 999 = present, -999 absent

// start state: real vector of length2 = start state in
// format (ESBL-, ESBL+) ie esbl positive coded as
// [0,1] and ESBL negative coded as [1,0]

// end state: integer length 1, final state.

```

```
// this will also generate and save log-likelihoods to do model comparison with loo.
```

```
functions {
```

```
// Time varying covariate value calculation
// Needs to be passed a 1d array of covariates
// each 3 entries are (cov_start_time, cov_end_time, prev_cov_end_time)
// prev_cov_end_time is coded as
// t of prev cov end time if has been exposure, pos no if not
// Needs to return a matrix with n_cov rows and 1 column
// to act on the alphas and betas of the model
```

```
// n_covs is an array with integer for each cov
// 1 = not time varying and coded with prev time- present if > 0
// and absent of < 0
// 2 = time varying but no decay; prev time is ignored
// 3 = time varying with decay. If there is no
// exposure in this block, set stop_time to < 0
```

```
real[] return_time_varying_coefs_exp_flat(
  real[] cov_mat_passed,
  real t1,
  int[] n_covs_passed,
  real[] gamma_passed
) {
  real out_vars[size(n_covs_passed)];
  int s;
  int f;
  int p;

  for (n in 1:size(n_covs_passed)) {
    s = 1 + ((n-1)*3);
    f = s + 1;
    p = f + 1;
    // for each row in cov matrix (ie each covariate)
    if (n_covs_passed[n] == 3) {
      // gamma decay
```

```

if (cov_mat_passed[f] > 0) { //if there is exposure this block
  if (t1 <= cov_mat_passed[f] && t1 >= cov_mat_passed[s]) {
    // if exposure is happening now
    // set value to 1
    out_vars[n] = 1;
  } else if (t1 > cov_mat_passed[f]) {
    // otherwise if there is exposure in this block
    // and this covariate is set to have a decaying effect
    // and time is after it has stopped
    // set value to decay from stop time
    out_vars[n] = exp((t1-cov_mat_passed[f])/(-1*gamma_passed[n]));
  } else if (t1 < cov_mat_passed[s] && cov_mat_passed[p] < 0) {
    // otherwise, if time is before start time
    // and there is previous exposure
    // set value to decay from previous time
    out_vars[n] = exp((t1-cov_mat_passed[p])/(-1*gamma_passed[n]));
  } else {
    // otherwise set to 0
    out_vars[n] = 0;
  }
} else { // if there is no exposure in this block
  if (cov_mat_passed[p] < 0) { // if there is previous exposure
    out_vars[n] = exp((t1-cov_mat_passed[p])/(-1*gamma_passed[n]));
  } else {
    out_vars[n] = 0;
  }
}

} else if (n_covs_passed[n] == 2) {

  if (t1 <= cov_mat_passed[f] && t1 >= cov_mat_passed[s]) {
    // if exposure is happening now
    // set value to 1
    out_vars[n] = 1;
  } else {
    out_vars[n] = 0;
  }
} else if (n_covs_passed[n] == 1) {
  if (cov_mat_passed[p] > 0) {

```

```

        out_vars[n] = 1;
    } else {
        out_vars[n] = 0;
    }
}
} // end of for loop
return out_vars;
} // end of fn

// function to return lambda(t) and mu(t)
// this should take a vector of length n_cov of time
// varying values of the covariates of the betas
// (from the time varying coef fn)
// and two vectors of length n_cov of parameters
// the alphas (that act on mu)
// and the betas (that act on lambda)
// and return a vect or of length two for the
// values of lambda(t) and mu(t)

// real[] return_time_var_transition_hazard(
//     real
// )

// differential state equation

real[] twostateODE2_flat(real t, // time
real[] y, // state
real[] theta, // parameters
real[] x_r, //data
int[] x_i) { // data

    // y is state as [p0,p1]
    // theta defined as
    // [ lambda, mu, gamma0, ... gamman,
    //  alpha0, alpha1, ... alphan,
    //  beta0 ... betan ]
    // where n is number of covariates
    // data x_r is 1d array of covariates, 3 for each covariate
    // x_i is array of covariate type as

```

```

// [number of non-timedep var,
//number of timedep nongamma var,
//number of gamma var,
// then an integer for each cov:
//1 (non timedep),2(nongamm) or 3(gamma)]

real dydt[2];
real coefs[size(x_i[])-3]; //vector of coefs
real alphaz[size(x_i[])-3]; // vector of alphas
real betaz[size(x_i[])-3]; // vector of betas
real gammaz[x_i[3]];
real lambda_pr;
real mu_pr;
real lambda0;
real mu0;
lambda0 = theta[1];
mu0 = theta[2];
gammaz = theta[3:(2+ x_i[3])];
alphaz = theta[(3+ x_i[3]):(3+x_i[3] + x_i[1] + x_i[2] + x_i[3] -1)] ;
betaz = theta[(3+x_i[3] + x_i[1] + x_i[2] +
x_i[3]):(2+x_i[3] + 2*(x_i[1] + x_i[2] + x_i[3]))];
coefs = return_time_varying_coefs_exp_flat(x_r, t, x_i[4:size(x_i)], gammaz);
lambda_pr = lambda0*exp(dot_product(coefs, betaz));
mu_pr = mu0*exp(dot_product(coefs, alphaz));

dydt[1] = -y[1]*lambda_pr + y[2]*mu_pr;
dydt[2] = y[1]*lambda_pr - y[2]*mu_pr;
return dydt;
} // end of function
} // end of block

data {
  int < lower = 1 > N; // Number of rows of data
  int <lower = 0> n_covs[3]; //[nontimevary, timevarynogamma, timevarygamma]
  int covs_type[sum(n_covs)]; // integer for each cov to define type
  real t[N]; // end time
  real cov_mat[N,sum(n_covs[])*3]; // array of covariates, 3 rows for each
  real start_state[N,2]; // start state (at t=0) in form [p0,p1]
  int end_state[N]; // end state (at t) as integer

```

```

}

transformed data {
  int x_i_pass[3 + sum(n_covs)];
  x_i_pass[] = append_array(n_covs[], covs_type[]);
}

parameters {
  real < lower = 0 > lambda;
  real < lower = 0 > mu;
  real <lower = 0> gammas[n_covs[3]];
  real alphas[sum(n_covs[])];
  real betas[sum(n_covs[])];
}

transformed parameters {
  real theta[2 + 2*(sum(n_covs)) + n_covs[3]];
  theta[1] = lambda;
  theta[2] = mu;
  theta[3:(2+ n_covs[3])] = gammas[];
  theta[(3+ n_covs[3]):(3+n_covs[3] + sum(n_covs) -1)] = alphas[];
  theta[(3+n_covs[3] + sum(n_covs)):(2+n_covs[3] + 2*(sum(n_covs)))] = betas[];
}

model {
  real temp[1,2];
  lambda ~ normal(0,0.2);
  mu ~ normal(0,0.2);
  alphas ~ normal(0,2);
  betas~ normal(0,2);
  gammas ~ normal(0,100);

  for (n in 1:N) {

    temp = integrate_ode_rk45(twostateODE2_flat,
      start_state[n],
      0, t[n:n],
      theta[],
      cov_mat[n],

```



```

        x_i_pass[], 1e-6,1e-6,1e6);

        if (end_state[n] == 1) {
            target += log(temp[1,2]);
        } else {
            target += log(temp[1,1]);
        }
    }
}

generated quantities {
    // needed for loo
    vector[N] log_lik;
    real temp[1,2];
    for (n in 1:N) {
        temp = integrate_ode_rk45(twostateODE2_flat,
            start_state[n],
            0,
            t[n:n],
            theta[],
            cov_mat[n],
            x_i_pass[],
            1e-6,1e-6,1e6);
        if (end_state[n] == 1) {
            log_lik[n] = log(temp[1,2]);
        } else {
            log_lik[n] = log(temp[1,1]);
        }
    }
}

```

```
// Stan model for msm style interval censored model, stepwise constant covariates
```

```
functions {

  // Differential state equations for solving

  real[] twostateODE(real t,          // time
    real[] y,          // state
    real[] theta,      // parameters
    real[] x_r,        // data (real)
    int[] x_i) {      // data (integer)

    real dydt[2];
    real lambda;
    real mu;
    real ab_alpha0;
    real ab_beta0;
    real hosp_alpha1;
    real hosp_beta1;

    real lambda_beta_sum;
    real mu_alpha_sum;

    lambda= theta[1] ;
    mu = theta[2];
    ab_alpha0 = theta[3];
    ab_beta0 = theta[4];
    hosp_alpha1 = theta[5];
    hosp_beta1 = theta[6];

    lambda_beta_sum = 0;
    mu_alpha_sum = 0;

    // first coef, abx, start x_r[1] and end time x_r[2]

    if (x_r[1] == 999) {
      // dont do anything, there is nothing for this covariate
    } else if (t <= x_r[2] && t >= x_r[1]) {
      lambda_beta_sum = lambda_beta_sum + ab_beta0;
```

```

mu_alpha_sum = mu_alpha_sum + ab_alpha0;
}

// second coef coef, abx, start x_r[3] and end time x_r[4]

if (x_r[3] == 999) {
// don't do anything, there is nothing for this covariate
} else if (t <= x_r[4] && t >= x_r[3]) {
lambda_beta_sum = lambda_beta_sum + hosp_beta1;
mu_alpha_sum = mu_alpha_sum + hosp_alpha1;
}

dydt[1] = -y[1]*lambda*exp(lambda_beta_sum) + y[2]*mu*exp(mu_alpha_sum);
dydt[2] = y[1]*lambda*exp(lambda_beta_sum) - y[2]*mu*exp(mu_alpha_sum);

return dydt;
}
}

data {
int < lower = 1 > N; // Sample size
real t[N]; // end time
real start_state[N,2]; // start state (at t_start) in form [p0,p1]
int end_state[N]; // end state (at t) as integer
real covariates[N,4]; // covariate start and end times
                        // (as ab_start, ab_end, hosp_start, hosp_end)
}

transformed data {
// real x_r[0];
int x_i[0];

}

parameters {
real < lower = 0 > lambda;
real < lower = 0 > mu;

```

```

real ab_alpha0;
real ab_beta0;
real hosp_alpha1;
real hosp_beta1;
// real < lower = 0 > gamma;
}

transformed parameters {
real theta[6];
theta[1] = lambda;
theta[2] = mu;
theta[3] = ab_alpha0;
theta[4] = ab_beta0;
theta[5] = hosp_alpha1;
theta[6] = hosp_beta1;

}

model {
real temp[1,2];
lambda ~ normal(0,0.2);
mu ~ normal(0,0.2);
ab_alpha0 ~ normal(0,2);
ab_beta0 ~ normal(0,2);
hosp_alpha1 ~ normal(0,2);
hosp_beta1 ~ normal(0,2);
//gamma ~ normal(20,20);
for (n in 1:N) {
temp = integrate_ode_rk45(twostateODE,
start_state[n],
0,
t[n:n],
theta,
covariates[n],
x_i,
1E-6,1E-6, 1E6);

if (end_state[n] == 1) {
target += log(temp[1,2]);
}
}
}

```

```

    } else {
    target += log(temp[1,1]);
    }
  }
}

generated quantities {
  vector[N] log_lik;
  real temp[1,2];
  for (n in 1:N) {
    temp = integrate_ode_rk45(twostateODE,
    start_state[n],
    0, t[n:n],
    theta,
    covariates[n],
    x_i,
    1E-6,1E-6, 1E6);

    if (end_state[n] == 1) {
      log_lik[n] = log(temp[1,2]);
    } else {
      log_lik[n] = log(temp[1,1]);
    }
  }
}

// The posterior predictive distribution

```



## Chapter 9

# Conclusions and further work

### 9.1 Introduction

In this thesis, I have presented the findings from a clinical study with two broad aims: first, to describe the presentation, aetiology, outcomes and determinants of outcomes in adults in Blantyre, Malawi; and second, to describe colonisation with ESBL-E as sepsis survivors pass from the hospital to the community, and an analysis to identify determinants of carriage. In Chapter one, I presented a hypothesis that it is possible to improve outcomes for patients with sepsis whilst reducing over-reliance on broad spectrum antibacterials such as ceftriaxone, and hence minimise pressure for the development of antimicrobial resistance, by rapidly initiating antimicrobial therapy to cover all pathogens likely to cause sepsis in Blantyre. In this chapter I review the findings of this thesis in the context of this hypothesis, and suggest directions for future work.

### 9.2 Summary of findings

The clinical study of sepsis presented in Chapters 3 and 4 confirm that sepsis in Malawi is very different from sepsis in high-income low-HIV and low-TB prevalence settings from where most studies of sepsis arise. Malawian patients presenting with sepsis, as elsewhere in sSA, are young, and predominantly HIV-infected. They have been sick for a longer period of time (median 7 days) with clinical syndromes/pathogens not typically associated with fulminant disease in high income settings i.e. typhoid fever, disseminated tuberculosis, and malaria. The easily modifiable determinants of sepsis outcome that have received so much attention in high-income settings - rapid administration of antibacterials and fluids - were not associated with survival in this cohort. This could be due to a true lack of effect, or to underpowering, but it highlights the need for data from sSA to guide the development of

sepsis protocols relevant to sSA.

TB was the commonest cause of sepsis identified in the cohort - 34% of participants had at least one positive diagnostic test for TB, rising to 50% in HIV-infected participants - and administration of antituberculous chemotherapy showed an association with survival (RR 1.22 for survival [95% CI 1.03-1.46] in propensity-score analysis). Confidence intervals are large, but the effect size for benefit in the anaemic (for example) is larger than that in those with confirmed TB, suggesting that there may be a role for empiric TB therapy in sepsis, but the role of TB therapy in the treatment of the critically unwell is unknown, and must be the focus of further work (see below). The 28-day mortality of sepsis in Blantyre was lower than expected at 18% (95% CI 13-23%), and lower than historical mortality estimates from across the continent described in Chapter 1. The reasons for this are speculative, but ART coverage was high compared to historical cohorts. Despite this, it seems likely that the presentation of sepsis in a number of participants in this study was a manifestation of ART failure. Few participants were switched to second-line therapy, and longer-term mortality (beyond 28 days) was significant in the HIV-infected with a near doubling of mortality from 19% (95% CI 13-26%) at 28 days to 36% (95% CI 28-45%) at 180 days. Together, these data suggest that a HIV test at admission may have a role in triage of initial antimicrobial therapy in sepsis in sSA and that urgent HIV viral load measurement in those with sepsis and suspicion of ART failure could help to facilitate rapid switch on to second line therapy.

In Chapter 5, I presented details of gut mucosal colonisation with ESBL-E in sepsis participants but also antibiotic-unexposed hospital inpatients and community members. Community carriage is common, with a baseline community carriage prevalence of 28% (95% CI 20-38%) and factors associated with baseline carriage (household crowding, unprotected water source and sample collection during rainy season) strongly suggest that within-household and environmental transmission routes are important. Co-trimoxazole preventative therapy (CPT) was also associated with ESBL-E carriage at baseline, as was recent hospitalisation. In antibacterial-exposed inpatients there is a dramatic rise in ESBL-E colonisation prevalence, which is not seen in the antibacterial-unexposed. As expected, broad spectrum antibacterial exposure, largely ceftriaxone, was near-universal in the sepsis cohort with a median 5 days (IQR 3-7) of exposure, but this is dwarfed in person-days of exposure by co-trimoxazole. This is perhaps unsurprising given that co-trimoxazole is mandated lifelong for all people living with HIV in Malawi, but the magnitude of the difference is more than an order of magnitude: 14,447 person days of exposure of co-trimoxazole versus 997 for ceftriaxone.

Genomic analyses using whole genome sequencing of a subset of cultured isolates reveals no obvious restrictions of mixing of *E. coli*, whether internationally or between hospital and community settings. The exception to this is a putative recently arrived high-risk clone, ST410, which shows some healthcare association. The genomic landscape of ESBL



in Blantyre is dominated by *bla<sub>CTX-M</sub>*, reflecting the global situation, and, worryingly, I describe (to my knowledge) the first carbapenemase identified in Enterobacteriaceae in Malawi, a *bla<sub>NDM-5</sub>* carried on a globally successful IncX plasmid. Other identified AMR genes reflect the local antibiotic pressures: in particular, co-trimoxazole resistance determinants are near-ubiquitous. Using WGS as a high-resolution typing tool by clustering bacteria core genes and ESBL-containing contigs I show that it is the ESBL contig-bacteria combination that is conserved within patients, suggesting that the unit of transmission in this system is the bacteria, rather than transfer of MGE. I also demonstrate significant turnover of ESBL-E: beyond 35 days, two samples from within a patient are no more likely to contain the same ESBL contig-bacteria combination than due to chance alone.

Finally, I develop and fit two-state continuous-time Markov models in a Bayesian framework to understand the drivers of ESBL-E carriage in this cohort. I find that hospitalisation increases both ESBL-E acquisition and loss with a net effect of rapid ESBL-E acquisition following hospitalisation. Antibacterial exposure acts to prolong carriage by reducing the rate of loss, and I find support in the data for a prolonged post-antibiotic effect with a long half-life of 43 days (95% CI 15-98 days) that acts to prolong carriage long after antibacterial exposure has finished. Simulations show that there is synergy between hospitalisation and antibacterial exposure which produces the sharp rises in ESBL-E colonisation prevalence, but also that the post-antibiotic effect results in significantly prolonged carriage. In addition, the post-antibiotic effect means that short courses of antibacterials have, in terms of ESBL-E carriage, similar effects to prolonged courses, with significant implications for antimicrobial stewardship. Co-trimoxazole preventative therapy seems to be a significant driver of long-term carriage.

### 9.3 Conclusions and future research priorities

In light of these findings, is it possible to provide suggestions for antibiotic strategies for sepsis in Blantyre that will improve outcomes whilst minimising acquisition of ESBL-E? I suggest not for severely unwell patients that urgently need antimicrobials: the post-antibiotic effect from the longitudinal models means that a short course of antibacterials (e.g. 2 days) has a similar effect to a longer one (e.g. more than 7 days), at least in terms of ESBL-E acquisition, and that all the considered antibacterials seem to have a similar effect of ESBL-E carriage. The models imply therefore that the antimicrobial stewardship intervention to best minimise ESBL-E carriage is to avoid antibacterials altogether. This is not feasible in these very unwell patients, in whom initial empiric antimicrobial therapy is certainly appropriate. However health systems interventions that support the safe restriction of antimicrobials in patients that do not require them such as patients with viral respiratory tract infections (e.g. improved point of care diagnostics) may achieve this effect. The data

I present therefore highlight several areas of further research priority.

### **1. Identification of TB and initiation of antituberculous chemotherapy at presentation**

There is a strong suggestion that antituberculous chemotherapy may improve outcomes in sepsis in Blantyre, but the best ways to deploy this therapy in the critically unwell are unknown. Whether empiric therapy is warranted depends on the diagnostic ability of the available tests for TB and, since performing the testing for urinary lipoarabinomannan (uLAM) that I have presented here, a more sensitive uLAM assay has become available. To truly make recommendation for the use of TB therapy in sepsis, it is necessary to understand the impact of this improved diagnostic on the analysis I have presented. To that end the stored urines from this study will be tested with the new, more sensitive fujiLAM assay. To understand the reasons for the unexpectedly low prevalence of TB bloodstream infection, these will be tested alongside Xpert Ultra testing of stored blood samples. These results will allow clear recommendations for the use of TB therapy in sepsis, which can be taken forward for testing in clinical trials. The pharmacokinetics of standard quadruple antituberculous therapy in the critically unwell are not, however, well described and pharmacokinetic studies in this population, perhaps as part of clinical trials of TB therapy in sepsis, should be undertaken. More broadly, interventions for disseminated TB have generally focussed on rapid diagnosis of TB, following the standard diagnostic and treatment paradigms for outpatient TB. I suggest that the data I present here provide support for a clinical trial of presumptive TB therapy in critically unwell inpatients.

### **2. Investigation of post 28-day mortality in HIV-infected participants**

The cause of the significant post-28 day mortality in HIV-infected participants warrants further scrutiny. This requires more in depth clinical follow up than was possible in this study, but a starting point is to clearly define whether the participants in this study who seem to be failing ART had true HIV virological failure. This can be achieved with retrospective testing of stored blood samples. The global success of ART roll out means that failure of first line therapy is likely to be increasingly common, and strategies to rapidly identify ART failure and switch people to second line may be needed. It may be that hospital admissions with severe infection provide such an opportunity. Demonstrating virologic failure is a necessary first step.

### **3. Understand the biology of post-antibiotic selection pressure for ESBL-E colonisation**

The key finding from the longitudinal modelling of ESBL-E carriage is that there is a prolonged post-antibiotic effect. It is this that leads me to state that the overall hypothesis of this thesis - that we can improve sepsis outcomes whilst minimising ESBL-E acquisition - cannot be supported with the available data, because a shorter course of antibacterials in

hospitalised patients is predicted to have a similar effect to a longer one. However, this also makes the post antibiotic effect an attractive target for intervention. I have demonstrated an association, but there is a need to understand the biology. How is it mediated? I suggest that a likely candidate, and a good starting point, is the composition of the gut microbiota. The effect of even short courses of antimicrobials on the microbiota can be profound, and antibacterial-mediated destruction of microbiota colonisation resistance to ESBL-E could certainly cause the effects I have described. Total stool DNA has been extracted in real time from the participants in the study, and shotgun metagenomic sequencing of these samples can be used to start to assess the role of the microbiota in mediating the post antibiotic effect: changes in the microbiota predictive of ESBL-E colonisation (or ESBL-E abundance) and identification of taxa associated with non-colonisation with ESBL-E at baseline could provide an understanding of mechanism. Ultimately, this could suggest microbiota-modulating therapies to promote colonisation resistance to ESBL-E.

Metagenomics can also address one of the limitations of the genomic analysis: that only one colony pick from each participant at each time point was sequenced, so the total within-host ESBL-E diversity is unknown. Metagenomics can not tell the whole story, however, as it can not place AMR genes within bacteria, but must be supplemented by further colony picks for whole genome sequencing. Stored plate sweeps from all the participants in the study can facilitate this. In addition, the work is already under way to sequence *E. coli* from all remaining samples and one *K. pneumoniae* from all samples in which this species was identified, using short-read sequencing. This will allow comparison of the AMR determinants carried within these two species within one host, over time, and allow an understanding of the extent to which horizontal gene transfer is occurring. To truly define the MGE upon which ESBL genes are carried requires long-read sequencing. This work is also under way on a number of representative isolates. This will also allow an analysis of the genomic environment of the apparently inactive *catB4* genes that were frequently seen in this analysis.

#### **4. Re-evaluate co-trimoxazole preventative therapy protocols in an era of rising Gram-negative resistance**

Finally, my analysis of determinants of ESBL-E carriage suggests that co-trimoxazole preventative therapy (CPT) may be a major driver of AMR carriage. The exact mechanism of mortality benefit of CPT is unknown and it may be that its benefit is driven by e.g. antimalarial activity; studies to address this hypothesis are under way (e.g. TSCQ trial ClinicalTrials.gov identifier NCT01650558). The fact that co-trimoxazole resistance is near ubiquitous in the ESBL *E. coli* in this study along with the modelling findings suggests that CPT may be having strong unintended selection pressure for ESBL-E. In an era of rising Gram-negative resistance I suggest that in some cases the benefits of this therapy may be outweighed by its risks and that perhaps a more nuanced approach to CPT usage

than lifelong treatment for all people living with HIV is required, which is still sufficiently pragmatic to be useful in sSA. This would require clinical studies examining the outcomes of strategies such as stopping CPT once well established on ART, and my data suggest that AMR endpoints (such as ESBL-E carriage) should be included in such trials.

# Appendix: R Session Information

To aid reproducibility, the version numbers of all packages used in the generation of this thesis are reproduced here using the R *sessionInfo()* command.

```
## R version 3.6.0 (2019-04-26)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Mojave 10.14.6
##
## Matrix products: default
## BLAS: /System/Library/Frameworks/Accelerate.framework/Versions/
A/Frameworks/vecLib.framework/Versions/A/libBLAS.dylib
## LAPACK:
/Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_GB.UTF-8/en_GB.UTF-8/en_GB.UTF-8/C/en_GB.UTF-8/en_GB.UTF-8
##
## attached base packages:
## [1] stats4 parallel grid stats graphics grDevices utils
## [8] datasets methods base
##
## other attached packages:
## [1] deSolve_1.24 loo_2.1.0 rstan_2.18.2
## [4] StanHeaders_2.18.1 IRanges_2.18.1 S4Vectors_0.22.0
## [7] BiocGenerics_0.30.0 gggenes_0.3.2 igraph_1.2.4.1
## [10] ggraph_1.0.2 viridis_0.5.1 viridisLite_0.3.0
## [13] cowplot_0.9.4 phytools_0.6-60 maps_3.3.0
## [16] ape_5.3 ggtree_1.16.1 MatchIt_3.0.2
## [19] wBoot_1.0.3 simpleboot_1.1-7 boot_1.3-22
## [22] glue_1.3.1 broom_0.5.2 epitools_0.5-10
## [25] factoextra_1.0.5 FactoMineR_1.41 brms_2.9.0
## [28] Rcpp_1.0.1 survminer_0.4.4 ggsci_2.9
```

```

## [31] eq5d_0.1.0 survival_2.44-1.1 lubridate_1.7.4
## [34] scales_1.0.0 RColorBrewer_1.1-2 eulerr_5.1.0
## [37] UpSetR_1.4.0 pheatmap_1.0.12 forcats_0.4.0
## [40] stringr_1.4.0 purrr_0.3.2 readr_1.3.1
## [43] tidyr_0.8.3.9000 tibble_2.1.2 tidyverse_1.2.1
## [46] rsvg_1.3 DiagrammeRsvg_0.1 DiagrammeR_1.0.1
## [49] maptools_0.9-5 sp_1.3-1 ggmap_3.0.0
## [52] lme4_1.1-21 tmap_2.2 ggpubr_0.2
## [55] magrittr_1.5 meta_4.9-5 metafor_2.1-0
## [58] Matrix_1.2-17 kableExtra_1.1.0 knitr_1.23
## [61] reshape2_1.4.3 ggplot2_3.1.1 dplyr_0.8.1
## [64] plyr_1.8.4 bookdown_0.11
##
## loaded via a namespace (and not attached):
## [1] coda_0.19-2 clusterGeneration_1.3.4
## [3] dygraphs_1.1.1.6 Rook_1.1-1
## [5] data.table_1.12.2 inline_0.3.15
## [7] generics_0.0.2 leaflet_2.0.2
## [9] callr_3.2.0 combinat_0.0-8
## [11] ggfittext_0.7.0 webshot_0.5.1
## [13] xml2_1.2.0 httpuv_1.5.1
## [15] assertthat_0.2.1 xfun_0.7
## [17] hms_0.4.2 bayesplot_1.7.0
## [19] satellite_1.0.1 evaluate_0.14
## [21] promises_1.0.1 readxl_1.3.1
## [23] km.ci_0.5-2 DBI_1.0.0
## [25] htmlwidgets_1.3 mapview_2.7.0
## [27] ellipsis_0.1.0.9000 tmaptools_2.0-1
## [29] crosstalk_1.0.0 backports_1.1.4
## [31] V8_2.2 markdown_0.9
## [33] influenceR_0.1.0 vctrs_0.1.0.9003
## [35] abind_1.4-5 withr_2.1.2
## [37] ggforce_0.2.2 packrat_0.5.0
## [39] rgdal_1.4-4 treeio_1.8.1
## [41] xts_0.11-2 prettyunits_1.0.2
## [43] mnormt_1.5-5 cluster_2.0.9
## [45] lazyeval_0.2.2 crayon_1.3.4
## [47] pkgconfig_2.0.2 labeling_0.3
## [49] units_0.6-3 tweenr_1.0.1

```

```
## [51] nlme_3.1-140 rlang_0.3.4.9003
## [53] miniUI_0.1.1.1 colourpicker_1.0
## [55] downloader_0.4 rgexf_0.15.3
## [57] modelr_0.1.4 dichromat_2.0-0
## [59] cellranger_1.1.0 polyclip_1.10-0
## [61] matrixStats_0.54.0 phangorn_2.5.3
## [63] raster_2.9-5 KMsurv_0.1-5
## [65] zoo_1.8-6 base64enc_0.1-3
## [67] ggribes_0.5.1 processx_3.3.1
## [69] png_0.1-7 rjson_0.2.20
## [71] bitops_1.0-6 KernSmooth_2.23-15
## [73] visNetwork_2.0.7 classInt_0.3-3
## [75] brew_1.0-6 jpeg_0.1-8
## [77] shinystan_2.5.0 leaps_3.0
## [79] threejs_0.3.1 compiler_3.6.0
## [81] rstantools_1.5.1 plotrix_3.7-5
## [83] cli_1.1.0 ps_1.3.0
## [85] Brodningnag_1.2-6 mgcv_1.8-28
## [87] MASS_7.3-51.4 tidyselect_0.2.5
## [89] stringi_1.4.3 highr_0.8
## [91] yaml_2.2.0 ggrepel_0.8.1
## [93] bridgesampling_0.6-0 survMisc_0.5.5
## [95] fastmatch_1.1-0 tools_3.6.0
## [97] animation_2.6 RgoogleMaps_1.4.3
## [99] rstudioapi_0.10 foreign_0.8-71
## [101] gridExtra_2.3 farver_1.1.0
## [103] scatterplot3d_0.3-41 digest_0.6.19
## [105] rvcheck_0.1.3 rgeos_0.4-3
## [107] shiny_1.3.2 cmprsk_2.2-8
## [109] quadprog_1.5-7 lwgeom_0.1-7
## [111] later_0.8.0 httr_1.4.0
## [113] rsconnect_0.8.13 sf_0.7-4
## [115] colorspace_1.4-1 polylabelr_0.1.0
## [117] rvest_0.3.4 XML_3.98-1.19
## [119] splines_3.6.0 expm_0.999-4
## [121] tidytree_0.2.4 shinythemes_1.1.2
## [123] xtable_1.8-4 jsonlite_1.6
## [125] nloptr_1.2.1 flashClust_1.01-2
## [127] zeallot_0.1.0 R6_2.4.0
```

```
## [129] pillar_1.4.1 htmltools_0.3.6
## [131] mime_0.6 minqa_1.2.4
## [133] DT_0.7 class_7.3-15
## [135] codetools_0.2-16 pkgbuild_1.0.3
## [137] mvtnorm_1.0-10 lattice_0.20-38
## [139] numDeriv_2016.8-1 curl_3.3
## [141] gtools_3.8.1 shinyjs_1.0
## [143] rmarkdown_1.13 munsell_0.5.0
## [145] e1071_1.7-1 haven_2.1.0
## [147] gtable_0.3.0
```



# References

- 1 Majno G. The Ancient Riddle of Sepsis. *Journal of Infectious Diseases* 1991;**163**:937–45. doi:10.1093/infdis/163.5.937
- 2 Singer M, Deutschman CS, Seymour CW *et al.* The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 2016;**315**:801. doi:10.1001/jama.2016.0287
- 3 Poll T van der, Veerdonk FL van de, Scicluna BP *et al.* The immunopathology of sepsis and potential therapeutic targets. *Nature reviews Immunology* 2017;**17**:407–20. doi:10.1038/nri.2017.36
- 4 Fleischmann C, Scherag A, Adhikari NKJ *et al.* Assessment of Global Incidence and Mortality of Hospital-treated Sepsis. Current Estimates and Limitations. *American Journal of Respiratory and Critical Care Medicine* 2016;**193**:259–72. doi:10.1164/rccm.201504-0781OC
- 5 Kaukonen K-M, Bailey M, Suzuki S *et al.* Mortality related to severe sepsis and septic shock among critically ill patients in Australia and New Zealand, 2000-2012. *JAMA* 2014;**311**:1308–16. doi:10.1001/jama.2014.2637
- 6 Kahn JM, Davis BS, Yabes JG *et al.* Association Between State-Mandated Protocolized Sepsis Care and In-hospital Mortality Among Adults With Sepsis. *JAMA* 2019;**322**:240. doi:10.1001/jama.2019.9021
- 7 Seymour CW, Gesten F, Prescott HC *et al.* Time to Treatment and Mortality during Mandated Emergency Care for Sepsis. *New England Journal of Medicine* 2017;**376**:2235–44. doi:10.1056/NEJMoa1703058
- 8 Lewis JM, Feasey NA, Rylance J. Aetiology and outcomes of sepsis in adults in sub-Saharan Africa: a systematic review and meta-analysis. *Critical Care* 2019;**23**:212. doi:10.1186/s13054-019-2501-y
- 9 Andrews B, Semler MW, Muchemwa L *et al.* Effect of an Early Resuscitation Protocol on In-hospital Mortality Among Adults With Sepsis and Hypotension: A Randomized Clinical Trial. *Jama* 2017;**318**:1233–40. doi:10.1001/jama.2017.10913

- 10 Maitland K, Kiguli S, Opoka RO *et al.* Mortality after fluid bolus in African children with severe infection. *The New England journal of medicine* 2011;**364**:2483–95. doi:10.1056/NEJMoa1101549
- 11 Moore CC, Jacob ST, Banura P *et al.* Etiology of Sepsis in Uganda using a Quantitative PCR-based TaqMan Array Card. *Clin Infect Dis* 2019;**68**:266–72. doi:10.1093/cid/ciy472
- 12 Prasad N, Murdoch DR, Reyburn H *et al.* Etiology of Severe Febrile Illness in Low- and Middle-Income Countries: A Systematic Review. *PloS one* 2015;**10**:e0127962. doi:10.1371/journal.pone.0127962
- 13 Rubach MP, Maro VP, Bartlett JA *et al.* Etiologies of illness among patients meeting integrated management of adolescent and adult illness district clinician manual criteria for severe infections in northern Tanzania: implications for empiric antimicrobial therapy. *Am J Trop Med Hyg* 2015;**92**:454–62. doi:10.4269/ajtmh.14-0496
- 14 Crump JA, Morrissey AB, Nicholson WL *et al.* Etiology of Severe Non-malaria Febrile Illness in Northern Tanzania: A Prospective Cohort Study. *PLoS Neglected Tropical Diseases* 2013;**7**:e2324. doi:10.1371/journal.pntd.0002324
- 15 Musicha P, Cornick JE, Bar-Zeev N *et al.* Trends in antimicrobial resistance in blood-stream infection isolates at a large urban hospital in Malawi (1998–2016): a surveillance study. *The Lancet Infectious Diseases* 2017;**17**:1042–52. doi:10.1016/S1473-3099(17)30394-8
- 16 Iroh Tam P-Y, Musicha P, Kawaza K *et al.* Emerging Resistance to Empiric Antimicrobial Regimens for Pediatric Bloodstream Infections in Malawi (1998–2017). *Clinical Infectious Diseases* 2019;**69**:61–8. doi:10.1093/cid/ciy834
- 17 Jacob ST, Moore CC, Banura P *et al.* Severe sepsis in two Ugandan hospitals: a prospective observational study of management and outcomes in a predominantly HIV-1 infected population. *PLoS One* 2009;**4**:e7782. doi:10.1371/journal.pone.0007782
- 18 Waitt PI, Mukaka M, Goodson P *et al.* Sepsis carries a high mortality among hospitalised adults in Malawi in the era of antiretroviral therapy scale-up: a longitudinal cohort study. *The Journal of infection* 2015;**70**:11–9. doi:10.1016/j.jinf.2014.07.004
- 19 Ssekitoleko R, Jacob ST, Banura P *et al.* Hypoglycemia at admission is associated with in-hospital mortality in Ugandan patients with severe sepsis. *Crit Care Med* 2011;**39**:2271–6. doi:10.1097/CCM.0b013e3182227bd2
- 20 Ssekitoleko R, Pinkerton R, Muhindo R *et al.* Aggregate evaluable organ dysfunction predicts in-hospital mortality from sepsis in Uganda. *Am J Trop Med Hyg* 2011;**85**:697–702. doi:10.4269/ajtmh.2011.10-0692
- 21 Chimese SM, Andrews B, Lakhi S. The Etiology And Outcome Of Adult Patients Presenting With Sepsis To The University Teaching Hospital, Lusaka, Zambia. *Med*

*J Zambia* 2012;**39**:19–22. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5663186/pdf/nihms524771.pdf>

22 Moore CC, Hazard R, Saulters KJ *et al.* Derivation and validation of a universal vital assessment (UVA) score: a tool for predicting mortality in adult hospitalised patients in sub-Saharan Africa. *BMJ Glob Health* 2017;**2**:e000344. doi:10.1136/bmjgh-2017-000344

23 Huson MAM, Kalkman R, Stolp SM *et al.* The impact of HIV on presentation and outcome of bacterial sepsis and other causes of acute febrile illness in Gabon. *Infection* 2015;**43**:443–51. doi:10.1007/s15010-015-0753-2

24 Amir A, Saulters KJ, Muhindo R *et al.* Outcomes of patients with severe infection in Uganda according to adherence to the World Health Organization’s Integrated Management of Adolescent and Adult Illness fluid resuscitation guidelines. *J Crit Care* 2017;**41**:24–8. doi:10.1016/j.jcrc.2017.04.042

25 Andrews B, Muchemwa L, Kelly P *et al.* Simplified severe sepsis protocol: a randomized controlled trial of modified early goal-directed therapy in Zambia. *Crit Care Med* 2014;**42**:2315–24. doi:10.1097/ccm.0000000000000541

26 Andrews B, Semler MW, Muchemwa L *et al.* Effect of an Early Resuscitation Protocol on In-hospital Mortality Among Adults With Sepsis and Hypotension. *JAMA* 2017;**318**:1233. doi:10.1001/jama.2017.10913

27 Jacob ST, Banura P, Baeten JM *et al.* The impact of early monitored management on survival in hospitalized adult Ugandan patients with severe sepsis: a prospective intervention study. *Crit Care Med* 2012;**40**:2050–8. doi:10.1097/CCM.0b013e31824e65d7

28 Zida S, Kania D, Sotto A *et al.* Leptospirosis as Cause of Febrile Icteric Illness, Burkina Faso. *Emerging Infectious Diseases* 2018;**24**:1569–72. doi:10.3201/eid2408.170436

29 Guillebaud J, Bernardson B, Randriambolamanantsoa TH *et al.* Study on causes of fever in primary healthcare center uncovers pathogens of public health concern in Madagascar. *PLOS Neglected Tropical Diseases* 2018;**12**:e0006642. doi:10.1371/journal.pntd.0006642

30 Maze MJ, Cash-Goldwasser S, Rubach MP *et al.* Risk factors for human acute leptospirosis in northern Tanzania. *PLOS Neglected Tropical Diseases* 2018;**12**:e0006372. doi:10.1371/journal.pntd.0006372

31 Gadia CLB, Manirakiza A, Tekpa G *et al.* Identification of pathogens for differential diagnosis of fever with jaundice in the Central African Republic: a retrospective assessment, 2008–2010. *BMC Infectious Diseases* 2017;**17**:735. doi:10.1186/s12879-017-2840-8

32 Hagen RM, Frickmann H, Ehlers J *et al.* Presence of *Borrelia* spp. DNA in ticks, but absence of *Borrelia* spp. and of *Leptospira* spp. DNA in blood of fever patients in Madagascar. *Acta Tropica* 2018;**177**:127–34. doi:10.1016/j.actatropica.2017.10.002

- 33 Biscornet L, Dellagi K, Pagès F *et al.* Human leptospirosis in Seychelles: A prospective study confirms the heavy burden of the disease but suggests that rats are not the main reservoir. *PLOS Neglected Tropical Diseases* 2017;**11**:e0005831. doi:10.1371/journal.pntd.0005831
- 34 Dreyfus A, Dyal JW, Pearson R *et al.* Leptospira Seroprevalence and Risk Factors in Health Centre Patients in Hoima District, Western Uganda. *PLOS Neglected Tropical Diseases* 2016;**10**:e0004858. doi:10.1371/journal.pntd.0004858
- 35 Hercik C, Cosmas L, Mogeni OD *et al.* A diagnostic and epidemiologic investigation of acute febrile illness (AFI) in Kilombero, Tanzania. *PLOS ONE* 2017;**12**:e0189712. doi:10.1371/journal.pone.0189712
- 36 Chipwaza B, Mugasa JP, Selemani M *et al.* Dengue and Chikungunya Fever among Viral Diseases in Outpatient Febrile Children in Kilosa District Hospital, Tanzania. *PLoS Neglected Tropical Diseases* 2014;**8**:e3335. doi:10.1371/journal.pntd.0003335
- 37 Cash-Goldwasser S, Crump JA, Halliday JEB *et al.* Risk Factors for Human Brucellosis in Northern Tanzania. *The American Journal of Tropical Medicine and Hygiene* 2018;**98**:598–606. doi:10.4269/ajtmh.17-0125
- 38 Gafirita J, Njunwa KJ, Ruhirwa R *et al.* Seroprevalence of Brucellosis among Patients Attending a District Hospital in Rwanda. *The American Journal of Tropical Medicine and Hygiene* 2017;**97**:831–5. doi:10.4269/ajtmh.16-0632
- 39 Boone I, Henning K, Hilbert A *et al.* Are brucellosis, Q fever and melioidosis potential causes of febrile illness in Madagascar? *Acta Tropica* 2017;**172**:255–62. doi:10.1016/j.actatropica.2017.05.013
- 40 Glanville WA de, Conde-Álvarez R, Moriyón I *et al.* Poor performance of the rapid test for human brucellosis in health facilities in Kenya. *PLOS Neglected Tropical Diseases* 2017;**11**:e0005508. doi:10.1371/journal.pntd.0005508
- 41 Njeru J, Melzer F, Wareth G *et al.* Human Brucellosis in Febrile Patients Seeking Treatment at Remote Hospitals, Northeastern Kenya, 2014-2015. *Emerging infectious diseases* 2016;**22**:2160–4. doi:10.3201/eid2212.160285
- 42 Chipwaza B, Mhamphi GG, Ngatunga SD *et al.* Prevalence of Bacterial Febrile Illnesses in Children in Kilosa District, Tanzania. *PLOS Neglected Tropical Diseases* 2015;**9**:e0003750. doi:10.1371/journal.pntd.0003750
- 43 Feleke SM, Animut A, Belay M. Prevalence of Malaria among Acute Febrile Patients Clinically Suspected of Having Malaria in the Zeway Health Center, Ethiopia. *Japanese Journal of Infectious Diseases* 2015;**68**:55–9. doi:10.7883/yoken.JJID.2013.062
- 44 Amoako N, Duodu S, Dennis FE *et al.* Detection of Dengue Virus among Children

with Suspected Malaria, Accra, Ghana. *Emerging Infectious Diseases* 2018;**24**:1544–7. doi:10.3201/eid2408.180341

45 Njeru J, Henning K, Pletz MW *et al.* Febrile patients admitted to remote hospitals in Northeastern Kenya: seroprevalence, risk factors and a clinical prediction tool for Q-Fever. *BMC Infectious Diseases* 2016;**16**:244. doi:10.1186/s12879-016-1569-0

46 Mourembou G, Nzondo SM, Ndjoyi-Mbiguino A *et al.* Co-circulation of Plasmodium and Bacterial DNAs in Blood of Febrile and Afebrile Children from Urban and Rural Areas in Gabon. *The American Journal of Tropical Medicine and Hygiene* 2016;**95**:123–32. doi:10.4269/ajtmh.15-0751

47 Angelakis E, Mediannikov O, Socolovschi C *et al.* Coxiella burnetii-positive PCR in febrile patients in rural and urban Africa. *International Journal of Infectious Diseases* 2014;**28**:107–10. doi:10.1016/j.ijid.2014.05.029

48 Maina AN, Farris CM, Odhiambo A *et al.* Q Fever, Scrub Typhus, and Rickettsial Diseases in Children, Kenya, 2011–2012. *Emerging Infectious Diseases* 2016;**22**:883–6. doi:10.3201/eid2205.150953

49 Sothmann P, Keller C, Krumkamp R *et al.* Rickettsia felis Infection in Febrile Children, Ghana. *The American Journal of Tropical Medicine and Hygiene* 2017;**96**:16–0754. doi:10.4269/ajtmh.16-0754

50 Elfving K, Shakely D, Andersson M *et al.* Acute Uncomplicated Febrile Illness in Children Aged 2–59 months in Zanzibar – Aetiologies, Antibiotic Treatment and Outcome. *PLOS ONE* 2016;**11**:e0146054. doi:10.1371/journal.pone.0146054

51 Mourembou G, Lekana-Douki JB, Mediannikov O *et al.* Possible Role of Rickettsia felia in Acute Febrile Illness among Children in Gabon. *Emerging Infectious Diseases* 2015;**21**:1808–15. doi:10.3201/eid2110.141825

52 Vu DM, Mutai N, Heath CJ *et al.* Unrecognized Dengue Virus Infections in Children, Western Kenya, 2014–2015. *Emerging infectious diseases* 2017;**23**:1915–7. doi:10.3201/eid2311.170807

53 Waggoner J, Brichard J, Mutuku F *et al.* Malaria and Chikungunya Detected Using Molecular Diagnostics Among Febrile Kenyan Children. *Open Forum Infectious Diseases* 2017;**4**:ofx110. doi:10.1093/ofid/ofx110

54 Kolawole OM, Seriki AA, Irekeola AA *et al.* Dengue virus and malaria concurrent infection among febrile subjects within Ilorin metropolis, Nigeria. *Journal of Medical Virology* 2017;**89**:1347–53. doi:10.1002/jmv.24788

55 Nasir IA, Agbede OO, Dangana A *et al.* Dengue virus non-structural Protein-1 expression and associated risk factors among febrile Patients attending University of Abuja Teaching

Hospital, Nigeria. *Virus Research* 2017;**230**:7–12. doi:10.1016/j.virusres.2016.12.011

56 Ngoi CN, Price MA, Fields B *et al.* Dengue and Chikungunya Virus Infections among Young Febrile Adults Evaluated for Acute HIV-1 Infection in Coastal Kenya. *PLOS ONE* 2016;**11**:e0167508. doi:10.1371/journal.pone.0167508

57 Onoja A, Adeniji J, Olaleye O. High rate of unrecognized dengue virus infection in parts of the rainforest region of Nigeria. *Acta Tropica* 2016;**160**:39–43. doi:10.1016/j.actatropica.2016.04.007

58 Kajeguka DC, Kaaya RD, Mwakalinga S *et al.* Prevalence of dengue and chikungunya virus infections in north-eastern Tanzania: a cross sectional study among participants presenting with malaria-like symptoms. *BMC Infectious Diseases* 2016;**16**:183. doi:10.1186/s12879-016-1511-5

59 Sow A, Loucoubar C, Diallo D *et al.* Concurrent malaria and arbovirus infections in Kedougou, southeastern Senegal. *Malaria Journal* 2016;**15**:47. doi:10.1186/s12936-016-1100-5

60 Kayiwa JT, Nankya AM, Ataliba IJ *et al.* Confirmation of Zika virus infection through hospital-based sentinel surveillance of acute febrile illness in Uganda, 2014–2017. *Journal of General Virology* Published Online First: July 2018. doi:10.1099/jgv.0.001113

61 Makiala-Mandanda S, Ahuka-Mundeke S, Abbate JL *et al.* Identification of Dengue and Chikungunya Cases Among Suspected Cases of Yellow Fever in the Democratic Republic of the Congo. *Vector-Borne and Zoonotic Diseases* 2018;**18**:364–70. doi:10.1089/vbz.2017.2176

62 Muianga A, Pinto G, Massangaie M *et al.* Antibodies Against Chikungunya in Northern Mozambique During Dengue Outbreak, 2014. *Vector-Borne and Zoonotic Diseases* 2018;vbz.2017.2261. doi:10.1089/vbz.2017.2261

63 Mugabe VA, Ali S, Chelene I *et al.* Evidence for chikungunya and dengue transmission in Quelimane, Mozambique: Results from an investigation of a potential outbreak of chikungunya virus. *PloS one* 2018;**13**:e0192110. doi:10.1371/journal.pone.0192110

64 António VS, Muianga AF, Wieseler J *et al.* Seroepidemiology of Chikungunya Virus Among Febrile Patients in Eight Health Facilities in Central and Northern Mozambique, 2015–2016. *Vector-Borne and Zoonotic Diseases* 2018;**18**:311–6. doi:10.1089/vbz.2017.2227

65 Sow A, Faye O, Diallo M *et al.* Chikungunya Outbreak in Kedougou, Southeastern Senegal in 2009–2010. *Open Forum Infectious Diseases* 2018;**5**:ofx259. doi:10.1093/ofid/ofx259

66 Olajiga OM, Adesoye OE, Emilolorun AP *et al.* Chikungunya Virus Seroprevalence and Associated Factors among Hospital Attendees in Two States of Southwest Nigeria: A Preliminary Assessment. *Immunological Investigations* 2017;**46**:552–65. doi:10.1080/08820139.2017.1319383

- 67 Bahr NC, Sarosi GA, Meya DB *et al.* Seroprevalence of histoplasmosis in Kampala, Uganda. *Medical Mycology* 2016;**54**:295–300. doi:10.1093/mmy/myv081
- 68 Mandengue CE, Ngandjio A, Atangana PJ. Histoplasmosis in HIV-Infected Persons, Yaoundé, Cameroon. *Emerging Infectious Diseases* 2015;**21**:2094–6. doi:10.3201/eid2111.150278
- 69 Schwarzer G. meta: An R package for meta-analysis. *R News* 2007;**7**:40–5.
- 70 Bone RC, Balk RA, Cerra FB *et al.* Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. *Chest* 1992;**101**:1644–55.
- 71 Levy MM, Fink MP, Marshall JC *et al.* 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Critical Care Medicine* 2003;**31**:1250–6. doi:10.1097/01.CCM.0000050454.01978.3B
- 72 Vincent JL, Mendonça A de, Cantraine F *et al.* Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Working group on "sepsis-related problems" of the European Society of Intensive Care Medicine. *Critical care medicine* 1998;**26**:1793–800.
- 73 Seymour CW, Liu VX, Iwashyna TJ *et al.* Assessment of Clinical Criteria for Sepsis. *JAMA* 2016;**315**:762. doi:10.1001/jama.2016.0288
- 74 Shankar-Hari M, Phillips GS, Levy ML *et al.* Developing a New Definition and Assessing New Clinical Criteria for Septic Shock. *JAMA* 2016;**315**:775. doi:10.1001/jama.2016.0289
- 75 Jacob ST, West TE, Banura P. Fitting a square peg into a round hole: are the current Surviving Sepsis Campaign guidelines feasible for Africa? *Crit Care* 2011;**15**:117. doi:10.1186/cc9981
- 76 Prin M, Itaye T, Clark S *et al.* Critical Care in a Tertiary Hospital in Malawi. *World J Surg* 2016;**40**:2635–42. doi:10.1007/s00268-016-3578-y
- 77 Huson MA, Kalkman R, Grobusch MP *et al.* Predictive value of the qSOFA score in patients with suspected infection in a resource limited setting in Gabon. *Travel Med Infect Dis* 2017;**15**:76–7. doi:10.1016/j.tmaid.2016.10.014
- 78 Huson MAM, Katete C, Chunda L *et al.* Application of the qSOFA score to predict mortality in patients with suspected infection in a resource-limited setting in Malawi. *Infection* 2017;**45**:893–6. doi:10.1007/s15010-017-1057-5
- 79 Aluisio AR, Garbern S, Wiskel T *et al.* Mortality outcomes based on ED qSOFA score and HIV status in a developing low income country. *The American Journal of Emergency*

*Medicine* Published Online First: March 2018. doi:10.1016/j.ajem.2018.03.014

80 Rudd KE, Seymour CW, Aluisio AR *et al.* Association of the Quick Sequential (Sepsis-Related) Organ Failure Assessment (qSOFA) Score With Excess Hospital Mortality in Adults With Suspected Infection in Low- and Middle-Income Countries. *JAMA* 2018;**319**:2202. doi:10.1001/jama.2018.6229

81 Gaieski DF, Edwards JM, Kallan MJ *et al.* Benchmarking the Incidence and Mortality of Severe Sepsis in the United States\*. *Critical Care Medicine* 2013;**41**:1167–74. doi:10.1097/CCM.0b013e31827c09f8

82 Angus DC, Linde-Zwirble WT, Lidicker J *et al.* Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Critical care medicine* 2001;**29**:1303–10.

83 Tsertsvadze A, Royle P, Seedat F *et al.* Community-onset sepsis and its public health burden: a systematic review. *Systematic reviews* 2016;**5**:81. doi:10.1186/s13643-016-0243-3

84 Mayr FB, Yende S, Linde-Zwirble WT *et al.* Infection Rate and Acute Organ Dysfunction Risk as Explanations for Racial Differences in Severe Sepsis. *JAMA* 2010;**303**:2495. doi:10.1001/jama.2010.851

85 PRISM Investigators, Rowan KM, Angus DC *et al.* Early, Goal-Directed Therapy for Septic Shock — A Patient-Level Meta-Analysis. *New England Journal of Medicine* 2017;**376**:2223–34. doi:10.1056/NEJMoa1701380

86 Rangel-Frausto MS, Pittet D, Costigan M *et al.* The Natural History of the Systemic Inflammatory Response Syndrome (SIRS). *JAMA* 1995;**273**:117. doi:10.1001/jama.1995.03520260039030

87 Serafim R, Gomes JA, Salluh J *et al.* A Comparison of the Quick-SOFA and Systemic Inflammatory Response Syndrome Criteria for the Diagnosis of Sepsis and Prediction of Mortality. *Chest* 2018;**153**:646–55. doi:10.1016/j.chest.2017.12.015

88 Stevenson EK, Rubenstein AR, Radin GT *et al.* Two Decades of Mortality Trends Among Patients With Severe Sepsis. *Critical Care Medicine* 2014;**42**:625–31. doi:10.1097/CCM.0000000000000026

89 Winters BD, Eberlein M, Leung J *et al.* Long-term mortality and quality of life in sepsis: A systematic review\*. *Critical Care Medicine* 2010;**38**:1276–83. doi:10.1097/CCM.0b013e3181d8cc1d

90 Shah FA, Pike F, Alvarez K *et al.* Bidirectional Relationship between Cognitive Function and Pneumonia. *American Journal of Respiratory and Critical Care Medicine* 2013;**188**:586–92. doi:10.1164/rccm.201212-2154OC



- 91 Iwashyna TJ, Ely EW, Smith DM *et al.* Long-term Cognitive Impairment and Functional Disability Among Survivors of Severe Sepsis. *JAMA* 2010;**304**:1787. doi:10.1001/jama.2010.1553
- 92 Yende S, Linde-Zwirble W, Mayr F *et al.* Risk of cardiovascular events in survivors of severe sepsis. *American journal of respiratory and critical care medicine* 2014;**189**:1065–74. doi:10.1164/rccm.201307-1321OC
- 93 Bergh C, Fall K, Udumyan R *et al.* Severe infections and subsequent delayed cardiovascular disease. *European journal of preventive cardiology* 2017;**24**:1958–66. doi:10.1177/2047487317724009
- 94 Ou S-M, Chu H, Chao P-W *et al.* Long-Term Mortality and Major Adverse Cardiovascular Events in Sepsis Survivors. A Nationwide Population-based Study. *American Journal of Respiratory and Critical Care Medicine* 2016;**194**:209–17. doi:10.1164/rccm.201510-2023OC
- 95 Vugia DJ, Kiehlbauch JA, Yeboue K *et al.* Pathogens and predictors of fatal septicemia associated with human immunodeficiency virus infection in Ivory Coast, west Africa. *J Infect Dis* 1993;**168**:564–70.
- 96 Archibald LK, Dulk MO den, Pallangyo KJ *et al.* Fatal Mycobacterium tuberculosis bloodstream infections in febrile hospitalized adults in Dar es Salaam, Tanzania. *Clin Infect Dis* 1998;**26**:290–6.
- 97 Ssali FN. A prospective study of community-Acquired Bloodstream infections Among febrile Adults Admitted to Mulago Hospital. *Journal of acquired immune deficiency syndromes and human retrovirology* 1998;**19**:484–9.
- 98 McDonald LC, Archibald LK, Rheanpumikankit S *et al.* Unrecognised Mycobacterium tuberculosis bacteraemia among hospital inpatients in less developed countries. *Lancet* 1999;**354**:1159–63.
- 99 Archibald LK, McDonald LC, Nwanyanwu O *et al.* A hospital-based prevalence survey of bloodstream infections in febrile patients in Malawi: implications for diagnosis and therapy. *J Infect Dis* 2000;**181**:1414–20. doi:10.1086/315367
- 100 Bell M, Archibald LK, Nwanyanwu O *et al.* Seasonal variation in the etiology of bloodstream infections in a febrile inpatient population in a developing country. *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases* 2001;**5**:63–9. doi:10.1016/S1201-9712(01)90027-X
- 101 Lewis DK, Peters RPH, Schijffelen MJ *et al.* Clinical indicators of mycobacteraemia in adults admitted to hospital in Blantyre, Malawi. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* 2002;**6**:1067–74.

- 102 Gilks CF, Brindle RJ, Mwachari C *et al.* Disseminated Mycobacterium avium infection among HIV-infected patients in Kenya. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 1995;**8**:195–8.
- 103 Gottberg A von, Sacks L, Machala S *et al.* Utility of blood cultures and incidence of mycobacteremia in patients with suspected tuberculosis in a South African infectious disease referral hospital. *Int J Tuberc Lung Dis* 2001;**5**:80–6.
- 104 Munseri PJ, Talbot E a, Bakari M *et al.* The bacteraemia of disseminated tuberculosis among HIV-infected patients with prolonged fever in Tanzania. *Scandinavian journal of infectious diseases* 2011;**43**:696–701. doi:10.3109/00365548.2011.577802
- 105 Nakiyingi L, Ssengooba W, Nakanjako D *et al.* Predictors and outcomes of mycobacteremia among HIV-infected smear- negative presumptive tuberculosis patients in Uganda. *BMC Infectious Diseases* 2015;**15**. doi:10.1186/s12879-015-0812-4
- 106 Feasey NA, Banada PP, Howson W *et al.* Evaluation of Xpert MTB/RIF for detection of tuberculosis from blood samples of HIV-infected adults confirms Mycobacterium tuberculosis bacteremia as an indicator of poor prognosis. *J Clin Microbiol* 2013;**51**:2311–6. doi:10.1128/jcm.00330-13
- 107 Lawn SD, Kerkhoff AD, Burton R *et al.* Rapid microbiological screening for tuberculosis in HIV-positive patients on the first day of acute hospital admission by systematic testing of urine samples using Xpert MTB/RIF: a prospective cohort in South Africa. *BMC medicine* 2015;**13**:192. doi:10.1186/s12916-015-0432-2
- 108 Waddell RD, Lishimpi K, Reyn CF von *et al.* Bacteremia due to Mycobacterium tuberculosis or M. bovis, Bacille Calmette-Guerin (BCG) among HIV- positive children and adults in Zambia. *Aids* 2001;**15**:55–60.
- 109 Gupta-Wright A, Corbett EL, Oosterhout JJ van *et al.* Rapid urine-based screening for tuberculosis in HIV-positive patients admitted to hospital in Africa (STAMP): a pragmatic, multicentre, parallel-group, double-blind, randomised controlled trial. *Lancet (London, England)* 2018;**392**:292–301. doi:10.1016/S0140-6736(18)31267-4
- 110 Gupta RK, Lucas SB, Fielding KL *et al.* Prevalence of tuberculosis in post-mortem studies of HIV-infected adults and children in resource-limited settings. *AIDS* 2015;**29**:1987–2002. doi:10.1097/QAD.0000000000000802
- 111 World Health Organisation. Improving the diagnosis and treatment of smear-negative pulmonary and extrapulmonary tuberculosis among adults and adolescents: recommendations for HIV-prevalent and resource-constrained settings. 2007.
- 112 Holtz TH, Kabera G, Mthiyane T *et al.* Use of a WHO-recommended algorithm to reduce mortality in seriously ill patients with HIV infection and smear-negative pulmonary

tuberculosis in South Africa: an observational cohort study. *The Lancet Infectious diseases* 2011;**11**:533–40. doi:10.1016/S1473-3099(11)70057-3

113 Osawa R, Singh N. Cytomegalovirus infection in critically ill patients: a systematic review. *Critical Care* 2009;**13**:R68. doi:10.1186/cc7875

114 World Health Organisation. IMAI district clinician manual: hospital care for adolescents and adults. 2011.

115 Rajasingham R, Smith RM, Park BJ *et al.* Global burden of disease of HIV-associated cryptococcal meningitis: an updated analysis. *The Lancet Infectious Diseases* 2017;**17**:873–81. doi:10.1016/S1473-3099(17)30243-8

116 Wasserman S, Engel ME, Mendelson M. Burden of pneumocystis pneumonia in HIV-infected adults in sub-Saharan Africa: protocol for a systematic review. *Syst Rev* 2013;**2**:112. doi:10.1186/2046-4053-2-112

117 Rhodes A, Evans LE, Alhazzani W *et al.* Surviving Sepsis Campaign. *Critical Care Medicine* 2017;**45**:486–552. doi:10.1097/CCM.0000000000002255

118 Dünser MW, Festic E, Dondorp A *et al.* Recommendations for sepsis management in resource-limited settings. *Intensive Care Medicine* 2012;**38**:557–74. doi:10.1007/s00134-012-2468-5

119 Mer M, Schultz MJ, Adhikari NK *et al.* Core elements of general supportive care for patients with sepsis and septic shock in resource-limited settings. *Intensive Care Medicine* 2017;**43**:1690–4. doi:10.1007/s00134-017-4831-z

120 Thwaites CL, Lundeg G, Dondorp AM *et al.* Recommendations for infection management in patients with sepsis and septic shock in resource-limited settings. *Intensive Care Medicine* 2016;**42**:2040–2. doi:10.1007/s00134-016-4415-3

121 Misango D, Pattnaik R, Baker T *et al.* Haemodynamic assessment and support in sepsis and septic shock in resource-limited settings. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 2017;**111**:483–9. doi:10.1093/trstmh/try007

122 Rivers E, Nguyen B, Havstad S *et al.* Early Goal-Directed Therapy in the Treatment of Severe Sepsis and Septic Shock. *New England Journal of Medicine* 2001;**345**:1368–77. doi:10.1056/NEJMoa010307

123 Yealy DM, Kellum JA, Huang DT *et al.* A randomized trial of protocol-based care for early septic shock. *The New England journal of medicine* 2014;**370**:1683–93. doi:10.1056/NEJMoa1401602

124 Peake SL, Delaney A, Bailey M *et al.* Goal-Directed Resuscitation for Patients with Early Septic Shock. *New England Journal of Medicine* 2014;**371**:141001063014008.

doi:10.1056/NEJMoa1404380

125 Mouncey PR, Osborn TM, Power GS *et al.* Trial of Early, Goal-Directed Resuscitation for Septic Shock. *New England Journal of Medicine* 2015;**372**:1503-1512. doi:10.1056/NEJMoa1500896

126 Kumar A, Roberts D, Wood KE *et al.* Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Critical care medicine* 2006;**34**:1589–96. doi:10.1097/01.CCM.0000217961.75225.E9

127 Sterling SA, Miller WR, Pryor J *et al.* The Impact of Timing of Antibiotics on Outcomes in Severe Sepsis and Septic Shock. *Critical Care Medicine* 2015;**43**:1907–15. doi:10.1097/CCM.0000000000001142

128 Paul M, Shani V, Muchtar E *et al.* Systematic review and meta-analysis of the efficacy of appropriate empiric antibiotic therapy for sepsis. *Antimicrobial agents and chemotherapy* 2010;**54**:4851–63. doi:10.1128/AAC.00627-10

129 Amir A, Saulters KJ, Olum S *et al.* Outcomes of patients with severe sepsis after the first 6 hours of resuscitation at a regional referral hospital in Uganda. *J Crit Care* 2016;**33**:78–83. doi:10.1016/j.jcrc.2016.01.023

130 Chalya PL, Mabula JB, Koy M *et al.* Typhoid intestinal perforations at a University teaching hospital in Northwestern Tanzania: A surgical experience of 104 cases in a resource-limited setting. *World journal of emergency surgery : WJES* 2012;**7**:4. doi:10.1186/1749-7922-7-4

131 Ferrer R, Artigas A, Suarez D *et al.* Effectiveness of Treatments for Severe Sepsis. *American Journal of Respiratory and Critical Care Medicine* 2009;**180**:861–6. doi:10.1164/rccm.200812-1912OC

132 Boyd JH, Forbes J, Nakada T-a *et al.* Fluid resuscitation in septic shock: A positive fluid balance and elevated central venous pressure are associated with increased mortality\*. *Critical Care Medicine* 2011;**39**:259–65. doi:10.1097/CCM.0b013e3181feeb15

133 Vincent J-L, Sakr Y, Sprung CL *et al.* Sepsis in European intensive care units: results of the SOAP study. *Critical care medicine* 2006;**34**:344–53.

134 Lee SJ, Ramar K, Park JG *et al.* Increased fluid administration in the first three hours of sepsis resuscitation is associated with reduced mortality: a retrospective cohort study. *Chest* 2014;**146**:908–15. doi:10.1378/chest.13-2702

135 Leisman DE, Goldman C, Doerfler ME *et al.* Patterns and Outcomes Associated With Timeliness of Initial Crystalloid Resuscitation in a Prospective Sepsis and Septic Shock Cohort. *Critical Care Medicine* 2017;**45**:1596–606. doi:10.1097/CCM.0000000000002574

- 136 Leisman DE, Doerfler ME, Schneider SM *et al.* Predictors, Prevalence, and Outcomes of Early Crystalloid Responsiveness Among Initially Hypotensive Patients With Sepsis and Septic Shock. *Critical Care Medicine* 2018;**46**:189–98. doi:10.1097/CCM.0000000000002834
- 137 Maitland K, George EC, Evans JA *et al.* Exploring mechanisms of excess mortality with early fluid resuscitation: insights from the FEAST trial. *BMC Medicine* 2013;**11**:68. doi:10.1186/1741-7015-11-68
- 138 Ambler RP. The Structure of beta-lactamases. *Philosophical Transactions of the Royal Society B: Biological Sciences* 1980;**289**:321–31. doi:10.1098/rstb.1980.0049
- 139 Bush K, Jacoby GA. Updated functional classification of beta-lactamases. *Antimicrobial agents and chemotherapy* 2010;**54**:969–76. doi:10.1128/AAC.01009-09
- 140 Paterson DL, Bonomo RA. Extended-spectrum beta-lactamases: a clinical update. *Clinical microbiology reviews* 2005;**18**:657–86. doi:10.1128/CMR.18.4.657-686.2005
- 141 Bradford PA. Extended-Spectrum-Lactamases in the 21st Century: Characterization, Epidemiology, and Detection of This Important Resistance Threat. *Clinical Microbiology and Infection* 2001;**14**:933–51. doi:10.1128/CMR.14.4.933-951.2001
- 142 ABRAHAM EP, CHAIN E. An Enzyme from Bacteria able to Destroy Penicillin. *Nature* 1940;**146**:837–7. doi:10.1038/146837a0
- 143 Datta N, Kontomichalou P. Penicillinase synthesis controlled by infectious R factors in Enterobacteriaceae. *Nature* 1965;**208**:239–41.
- 144 Knothe H, Shah P, Krcmery V *et al.* Transferable resistance to cefotaxime, cefoxitin, cefamandole and cefuroxime in clinical isolates of *Klebsiella pneumoniae* and *Serratia marcescens*. *Infection* 1983;**11**:315–7.
- 145 Kliebe C, Nies BA, Meyer JF *et al.* Evolution of plasmid-coded resistance to broad-spectrum cephalosporins. *Antimicrobial agents and chemotherapy* 1985;**28**:302–7.
- 146 Liakopoulos A, Mevius D, Ceccarelli D. A Review of SHV Extended-Spectrum  $\beta$ -Lactamases: Neglected Yet Ubiquitous. *Frontiers in microbiology* 2016;**7**:1374. doi:10.3389/fmicb.2016.01374
- 147 Sougakoff W, Goussard S, Gerbaud G *et al.* Plasmid-mediated resistance to third-generation cephalosporins caused by point mutations in TEM-type penicillinase genes. *Reviews of infectious diseases* 1988;**10**:879–84.
- 148 Gold HS, Moellering RC. Antimicrobial-Drug Resistance. *New England Journal of Medicine* 1996;**335**:1445–53. doi:10.1056/NEJM199611073351907
- 149 Philippon A, Labia R, Jacoby G. Extended-spectrum beta-lactamases. *Antimicrobial agents and chemotherapy* 1989;**33**:1131–6.

- 150 Cantón R, Coque TM. The CTX-M beta-lactamase pandemic. *Current opinion in microbiology* 2006;**9**:466–75. doi:10.1016/j.mib.2006.08.011
- 151 Livermore DM, Canton R, Gniadkowski M *et al.* CTX-M: changing the face of ESBLs in Europe. *Journal of Antimicrobial Chemotherapy* 2006;**59**:165–74. doi:10.1093/jac/dkl483
- 152 Ben-Ami R, Rodríguez-Baño J, Arslan H *et al.* A Multinational Survey of Risk Factors for Infection with Extended-Spectrum  $\beta$ -Lactamase-Producing Enterobacteriaceae in Nonhospitalized Patients. *Clinical Infectious Diseases* 2009;**49**:682–90. doi:10.1086/604713
- 153 Bauernfeind A, Grimm H, Schweighart S. A new plasmidic cefotaximase in a clinical isolate of *Escherichia coli*. *Infection* 1990;**18**:294–8.
- 154 Bonnet R. Growing group of extended-spectrum beta-lactamases: the CTX-M enzymes. *Antimicrobial agents and chemotherapy* 2004;**48**:1–14. doi:10.1128/AAC.48.1.1-14.2004
- 155 Bevan ER, Jones AM, Hawkey PM. Global epidemiology of CTX-M  $\beta$ -lactamases: temporal and geographical shifts in genotype. *Journal of Antimicrobial Chemotherapy* 2017;**72**:2145–55. doi:10.1093/jac/dkx146
- 156 Karanika S, Karantanos T, Arvanitis M *et al.* Fecal Colonization With Extended-spectrum Beta-lactamase-Producing Enterobacteriaceae and Risk Factors Among Healthy Individuals: A Systematic Review and Metaanalysis. *Clinical Infectious Diseases* 2016;**63**:310–8. doi:10.1093/cid/ciw283
- 157 Goossens H, Ferech M, Vander Stichele R *et al.* Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. *The Lancet* 2005;**365**:579–87. doi:10.1016/S0140-6736(05)17907-0
- 158 Lai C-C, Wang C-Y, Chu C-C *et al.* Correlation between antibiotic consumption and resistance of Gram-negative bacteria causing healthcare-associated infections at a university hospital in Taiwan from 2000 to 2009. *Journal of Antimicrobial Chemotherapy* 2011;**66**:1374–82. doi:10.1093/jac/dkr103
- 159 Alvarez-Uria G, Gandra S, Laxminarayan R. Poverty and prevalence of antimicrobial resistance in invasive isolates. *International Journal of Infectious Diseases* 2016;**52**:59–61. doi:10.1016/J.IJID.2016.09.026
- 160 Denis B, Lafaurie M, Donay J-L *et al.* Prevalence, risk factors, and impact on clinical outcome of extended-spectrum beta-lactamase-producing *Escherichia coli* bacteraemia: a five-year study. *International Journal of Infectious Diseases* 2015;**39**:1–6. doi:10.1016/J.IJID.2015.07.010
- 161 Gorrie CL, Mirceta M, Wick RR *et al.* Antimicrobial-Resistant *Klebsiella pneumoniae* Carriage and Infection in Specialized Geriatric Care Wards Linked to Acquisition in the Referring Hospital. *Clinical Infectious Diseases* 2018;**67**:161–70. doi:10.1093/cid/ciy027

- 162 Mirelis B, Navarro F, Miró E *et al.* Community Transmission of Extended-Spectrum  $\beta$ -Lactamase. *Emerging Infectious Diseases* 2003;**9**:1024–5. doi:10.3201/eid0908.030094
- 163 Franciczek R, Sobieszczańska B, Grabowski M *et al.* Occurrence of extended-spectrum beta-lactamases among *Escherichia coli* isolates from hospitalized and healthy children. *Folia microbiologica* 2003;**48**:243–7.
- 164 Woerther P-L, Burdet C, Chachaty E *et al.* Trends in human fecal carriage of extended-spectrum  $\beta$ -lactamases in the community: toward the globalization of CTX-M. *Clinical microbiology reviews* 2013;**26**:744–58. doi:10.1128/CMR.00023-13
- 165 McNulty CAM, Lecky DM, Xu-McCrae L *et al.* CTX-M ESBL-producing Enterobacteriaceae: estimated prevalence in adults in England in 2014. *The Journal of antimicrobial chemotherapy* 2018;**73**:1368–88. doi:10.1093/jac/dky007
- 166 Wielders C, Hoek A van, Hengeveld P *et al.* Extended-spectrum  $\beta$ -lactamase- and pAmpC-producing Enterobacteriaceae among the general population in a livestock-dense area. *Clinical Microbiology and Infection* 2017;**23**:120.e1–8. doi:10.1016/J.CMI.2016.10.013
- 167 Ny S, Löfmark S, Börjesson S *et al.* Community carriage of ESBL-producing *Escherichia coli* is associated with strains of low pathogenicity: a Swedish nationwide study. *Journal of Antimicrobial Chemotherapy* 2017;**72**:582–8. doi:10.1093/jac/dkw419
- 168 Valverde A, Coque TM, Sanchez-Moreno MP *et al.* Dramatic Increase in Prevalence of Fecal Carriage of Extended-Spectrum  $\beta$ -Lactamase-Producing Enterobacteriaceae during Nonoutbreak Situations in Spain. *Journal of Clinical Microbiology* 2004;**42**:4769–75. doi:10.1128/JCM.42.10.4769-4775.2004
- 169 Li B, Sun J-Y, Liu Q-Z *et al.* High prevalence of CTX-M  $\beta$ -lactamases in faecal *Escherichia coli* strains from healthy humans in Fuzhou, China. *Scandinavian Journal of Infectious Diseases* 2011;**43**:170–4. doi:10.3109/00365548.2010.538856
- 170 Babu R, Kumar A, Karim S *et al.* Faecal carriage rate of extended-spectrum  $\beta$ -lactamase-producing Enterobacteriaceae in hospitalised patients and healthy asymptomatic individuals coming for health check-up. *Journal of Global Antimicrobial Resistance* 2016;**6**:150–3. doi:10.1016/j.jgar.2016.05.007
- 171 Reuland EA, Al Naiemi N, Kaiser AM *et al.* Prevalence and risk factors for carriage of ESBL-producing Enterobacteriaceae in Amsterdam. *The Journal of antimicrobial chemotherapy* 2016;**71**:1076–82. doi:10.1093/jac/dkv441
- 172 Woerther P-L, Andreumont A, Kantele A. Travel-acquired ESBL-producing Enterobacteriaceae: impact of colonization at individual and community level. *Journal of travel medicine* 2017;**24**:S29–34. doi:10.1093/jtm/taw101
- 173 Luvsansharav U-O, Hirai I, Nakata A *et al.* Prevalence of and risk factors associated

- with faecal carriage of CTX-M  $\beta$ -lactamase-producing Enterobacteriaceae in rural Thai communities. *Journal of Antimicrobial Chemotherapy* 2012;**67**:1769–74. doi:10.1093/jac/dks118
- 174 March A, Aschbacher R, Dhanji H *et al.* Colonization of residents and staff of a long-term-care facility and adjacent acute-care hospital geriatric unit by multiresistant bacteria. *Clinical Microbiology and Infection* 2010;**16**:934–44. doi:10.1111/J.1469-0691.2009.03024.X
- 175 Valverde A, Grill F, Coque TM *et al.* High rate of intestinal colonization with extended-spectrum- $\beta$ -lactamase-producing organisms in household contacts of infected community patients. *Journal of clinical microbiology* 2008;**46**:2796–9. doi:10.1128/JCM.01008-08
- 176 Rodriguez-Bano J, Lopez-Cerero L, Navarro MD *et al.* Faecal carriage of extended-spectrum  $\beta$ -lactamase-producing *Escherichia coli*: prevalence, risk factors and molecular epidemiology. *Journal of Antimicrobial Chemotherapy* 2008;**62**:1142–9. doi:10.1093/jac/dkn293
- 177 Duijkeren E van, Wienders CCH, Dierikx CM *et al.* Long-term Carriage of Extended-Spectrum  $\beta$ -Lactamase-Producing *Escherichia coli* and *Klebsiella pneumoniae* in the General Population in The Netherlands. *Clinical Infectious Diseases* 2018;**66**:1368–76. doi:10.1093/cid/cix1015
- 178 Alsterlund R, Carlsson B, Gezelius L *et al.* Multiresistant CTX-M-15 ESBL-producing *Escherichia coli* in southern Sweden: Description of an outbreak. *Scandinavian Journal of Infectious Diseases* 2009;**41**:410–5. doi:10.1080/00365540902896079
- 179 Zahar J, Lanternier F, Mechai F *et al.* Duration of colonisation by Enterobacteriaceae producing extended-spectrum  $\beta$ -lactamase and risk factors for persistent faecal carriage. *Journal of Hospital Infection* 2010;**75**:76–8. doi:10.1016/j.jhin.2009.11.010
- 180 Lohr IH, Rettedal S, Natas OB *et al.* Long-term faecal carriage in infants and intra-household transmission of CTX-M-15-producing *Klebsiella pneumoniae* following a nosocomial outbreak. *Journal of Antimicrobial Chemotherapy* 2013;**68**:1043–8. doi:10.1093/jac/dks502
- 181 Arcilla MS, Hattem JM van, Haverkate MR *et al.* Import and spread of extended-spectrum  $\beta$ -lactamase-producing Enterobacteriaceae by international travellers (COMBAT study): a prospective, multicentre cohort study. *The Lancet Infectious Diseases* 2017;**17**:78–85. doi:10.1016/S1473-3099(16)30319-X
- 182 Teunis PFM, Evers EG, Hengeveld PD *et al.* Time to acquire and lose carrier-ship of ESBL/pAmpC producing *E. coli* in humans in the Netherlands. *PLOS ONE* 2018;**13**:e0193834. doi:10.1371/journal.pone.0193834
- 183 Lartigue M-F, Poirel L, Aubert D *et al.* In vitro analysis of ISEcp1B-mediated mobilization of naturally occurring beta-lactamase gene blaCTX-M of *Kluyvera ascorbata*. *Antimi-*



*crobial agents and chemotherapy* 2006;**50**:1282–6. doi:10.1128/AAC.50.4.1282-1286.2006

184 Cantón R, María González-Alba J, Galán JC *et al.* CTX-M enzymes: origin and diffusion. *Frontiers in microbiology* 2012;**3**:110. doi:10.3389/fmicb.2012.00110

185 Zhao W-H, Hu Z-Q. Epidemiology and genetics of CTX-M extended-spectrum  $\beta$ -lactamases in Gram-negative bacteria. *Critical reviews in microbiology* 2013;**39**:79–101. doi:10.3109/1040841X.2012.691460

186 Carattoli A. Resistance Plasmid Families in Enterobacteriaceae. *Antimicrobial Agents and Chemotherapy* 2009;**53**:2227–38. doi:10.1128/AAC.01707-08

187 Coque TM, Novais Â, Carattoli A *et al.* Dissemination of Clonally Related Escherichia coli Strains Expressing Extended-Spectrum  $\beta$ -Lactamase CTX-M-15. *Emerging Infectious Diseases* 2008;**14**:195–200. doi:10.3201/eid1402.070350

188 Nicolas-Chanoine M-H, Bertrand X, Madec J-Y. Escherichia coli ST131, an intriguing clonal group. *Clinical microbiology reviews* 2014;**27**:543–74. doi:10.1128/CMR.00125-13

189 Stoesser N, Sheppard AE, Pankhurst L *et al.* Evolutionary History of the Global Emergence of the Escherichia coli Epidemic Clone ST131. *mBio* 2016;**7**:e02162. doi:10.1128/mBio.02162-15

190 Petty NK, Ben Zakour NL, Stanton-Cook M *et al.* Global dissemination of a multidrug resistant Escherichia coli clone. *Proceedings of the National Academy of Sciences* 2014;**111**:5694–9. doi:10.1073/pnas.1322678111

191 Bates D, Mächler M, Bolker B *et al.* Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 2015;**67**:1–48. doi:10.18637/jss.v067.i01

192 Guiral E, Pons MJ, Vubil D *et al.* Epidemiology and molecular characterization of multidrug-resistant Escherichia coli isolates harboring blaCTX-M group 1 extended-spectrum beta-lactamases causing bacteremia and urinary tract infection in Manhica, Mozambique. *Infect Drug Resist* 2018;**11**:927–36. doi:10.2147/idr.s153601

193 Karppinen M, Bernardino L, Anjos ED *et al.* Etiology of Childhood Otorrhea in Luanda, Angola, and a Review of Otitis Media in African Children. *Pediatr Infect Dis J* Published Online First: 2018. doi:10.1097/inf.0000000000002254

194 Ibrahim Y, Sani Y, Saleh Q *et al.* Phenotypic Detection of Extended Spectrum Beta lactamase and Carbapenemase Co-producing Clinical Isolates from Two Tertiary Hospitals in Kano, North West Nigeria. *Ethiop J Health Sci* 2017;**27**:3–10.

195 Kassam NA, Damian DJ, Kajeguka D *et al.* Spectrum and antibiogram of bacteria isolated from patients presenting with infected wounds in a Tertiary Hospital, northern Tanzania. *BMC Res Notes* 2017;**10**:757. doi:10.1186/s13104-017-3092-9

- 196 Legese MH, Weldearegay GM, Asrat D. Extended-spectrum beta-lactamase- and carbapenemase-producing Enterobacteriaceae among Ethiopian children. *Infect Drug Resist* 2017;**10**:27–34. doi:10.2147/idr.s127177
- 197 Manyahi J, Moyo SJ, Tellevik MG *et al.* Detection of CTX-M-15 beta-lactamases in Enterobacteriaceae causing hospital- and community-acquired urinary tract infections as early as 2004, in Dar es Salaam, Tanzania. *BMC Infect Dis* 2017;**17**:282. doi:10.1186/s12879-017-2395-8
- 198 Sangare SA, Rondinaud E, Maataoui N *et al.* Very high prevalence of extended-spectrum beta-lactamase-producing Enterobacteriaceae in bacteriemic patients hospitalized in teaching hospitals in Bamako, Mali. *PLoS One* 2017;**12**. doi:10.1371/journal.pone.0172652
- 199 Abera B, Kibret M, Mulu W. Extended-Spectrum beta (beta)-Lactamases and Antibio-gram in Enterobacteriaceae from Clinical and Drinking Water Sources from Bahir Dar City, Ethiopia. *PLoS One* 2016;**11**:e0166519. doi:10.1371/journal.pone.0166519
- 200 Agyekum A, Fajardo-Lubian A, Ansong D *et al.* blaCTX-M-15 carried by IncF-type plasmids is the dominant ESBL gene in Escherichia coli and Klebsiella pneumoniae at a hospital in Ghana. *Diagn Microbiol Infect Dis* 2016;**84**:328–33. doi:10.1016/j.diagmicrobio.2015.12.010
- 201 Breurec S, Bouchiat C, Sire JM *et al.* High third-generation cephalosporin resistant Enterobacteriaceae prevalence rate among neonatal infections in Dakar, Senegal. *BMC Infect Dis* 2016;**16**. doi:10.1186/s12879-016-1935-y
- 202 Buys H, Muloiwa R, Bamford C *et al.* Klebsiella pneumoniae bloodstream infections at a South African children’s hospital 2006-2011, a cross-sectional study. *BMC Infect Dis* 2016;**16**:570. doi:10.1186/s12879-016-1919-y
- 203 Eibach D, Belmar Campos C, Krumkamp R *et al.* Extended spectrum beta-lactamase producing Enterobacteriaceae causing bloodstream infections in rural Ghana, 2007-2012. *Int J Med Microbiol* 2016;**306**:249–54. doi:10.1016/j.ijmm.2016.05.006
- 204 Kpoda DS, Ajayi A, Somda M *et al.* Distribution of resistance genes encoding ESBLs in Enterobacteriaceae isolated from biological samples in health centers in Ouagadougou, Burkina Faso. *BMC Res Notes* 2018;**11**:471. doi:10.1186/s13104-018-3581-5
- 205 Kabwe M, Tembo J, Chilukutu L *et al.* Etiology, antibiotic resistance and risk factors for neonatal sepsis in a large referral center in Zambia. *Pediatric Infectious Disease Journal* 2016;**35**:e191–8. doi:10.1097/INF.0000000000001154
- 206 Leski TA, Taitt CR, Bangura U *et al.* High prevalence of multidrug resistant Enterobacteriaceae isolated from outpatient urine samples but not the hospital environment in Bo, Sierra Leone. *BMC Infect Dis* 2016;**16**:167. doi:10.1186/s12879-016-1495-1

- 207 Mohammed Y, Gadzama GB, Zailani SB *et al.* Characterization of Extended-Spectrum Beta-lactamase from *Escherichia coli* and *Klebsiella* Species from North Eastern Nigeria. *J Clin Diagn Res* 2016;**10**:Dc07–10. doi:10.7860/jcdr/2016/16330.7254
- 208 Naas T, Cuzon G, Robinson AL *et al.* Neonatal infections with multidrug-resistant ESBL-producing *E. cloacae* and *K. pneumoniae* in Neonatal Units of two different Hospitals in Antananarivo, Madagascar. *BMC Infect Dis* 2016;**16**:275. doi:10.1186/s12879-016-1580-5
- 209 Ndir A, Diop A, Faye PM *et al.* Epidemiology and Burden of Bloodstream Infections Caused by Extended-Spectrum Beta-Lactamase Producing Enterobacteriaceae in a Pediatric Hospital in Senegal. *PLoS One* 2016;**11**:e0143729. doi:10.1371/journal.pone.0143729
- 210 Ouedraogo AS, Sanou M, Kissou A *et al.* High prevalence of extended-spectrum ss-lactamase producing enterobacteriaceae among clinical isolates in Burkina Faso. *BMC Infect Dis* 2016;**16**:326. doi:10.1186/s12879-016-1655-3
- 211 Sangare SA, Maiga AI, Guindo I *et al.* Prevalence of ESBL-producing Enterobacteriaceae isolated from blood cultures in Mali. *J Infect Dev Ctries* 2016;**10**:1059–64. doi:10.3855/jidc.7536
- 212 Seni J, Falgenhauer L, Simeo N *et al.* Multiple ESBL-Producing *Escherichia coli* Sequence Types Carrying Quinolone and Aminoglycoside Resistance Genes Circulating in Companion and Domestic Farm Animals in Mwanza, Tanzania, Harbor Commonly Occurring Plasmids. *Front Microbiol* 2016;**7**:142. doi:10.3389/fmicb.2016.00142
- 213 Dramowski A, Cotton MF, Rabie H *et al.* Trends in paediatric bloodstream infections at a South African referral hospital. *BMC Pediatr* 2015;**15**:33. doi:10.1186/s12887-015-0354-3
- 214 Ireng LM, Kabego L, Kinunu FB *et al.* Antimicrobial resistance of bacteria isolated from patients with bloodstream infections at a tertiary care hospital in the Democratic Republic of the Congo. *S Afr Med J* 2015;**105**:752–5. doi:10.7196/SAMJnew.7937
- 215 Onanuga A, Omeje MC, Eboh DD. CARRIAGE OF MULTI-DRUG RESISTANT UROBACTERIA BY ASYMPTOMATIC PREGNANT WOMEN IN YENAGOA, BAYELSA STATE, NIGERIA. *Afr J Infect Dis* 2018;**12**:14–20. doi:10.21010/ajid.v12i2.3
- 216 Kateregga JN, Kantume R, Atuhair C *et al.* Phenotypic expression and prevalence of ESBL-producing Enterobacteriaceae in samples collected from patients in various wards of Mulago Hospital, Uganda. *BMC Pharmacol Toxicol* 2015;**16**:14. doi:10.1186/s40360-015-0013-1
- 217 Opintan JA, Newman MJ, Arhin RE *et al.* Laboratory-based nationwide surveillance of antimicrobial resistance in Ghana. *Infect Drug Resist* 2015;**8**:379–89. doi:10.2147/idr.s88725
- 218 Pons MJ, Vubil D, Guiral E *et al.* Characterisation of extended-spectrum beta-lactamases among *Klebsiella pneumoniae* isolates causing bacteraemia and

urinary tract infection in Mozambique. *J Glob Antimicrob Resist* 2015;**3**:19–25. doi:10.1016/j.jgar.2015.01.004

219 Rafaï C, Frank T, Manirakiza A *et al.* Dissemination of IncF-type plasmids in multiresistant CTX-M-15-producing Enterobacteriaceae isolates from surgical-site infections in Bangui, Central African Republic. *BMC Microbiol* 2015;**15**. doi:10.1186/s12866-015-0348-1

220 Adeyankinnu FA, Motayo BO, Akinduti A *et al.* A Multicenter Study of Beta-Lactamase Resistant *Escherichia coli* and *Klebsiella pneumoniae* Reveals High Level Chromosome Mediated Extended Spectrum beta Lactamase Resistance in Ogun State, Nigeria. *Interdiscip Perspect Infect Dis* 2014;**2014**:819896. doi:10.1155/2014/819896

221 Irengé LM, Kabego L, Vandenberg O *et al.* Antimicrobial resistance in urinary isolates from inpatients and outpatients at a tertiary care hospital in South-Kivu Province (Democratic Republic of Congo). *BMC Res Notes* 2014;**7**:374. doi:10.1186/1756-0500-7-374

222 Scherbaum M, Kusters K, Murbeth RE *et al.* Incidence, pathogens and resistance patterns of nosocomial infections at a rural hospital in Gabon. *BMC Infect Dis* 2014;**14**:124. doi:10.1186/1471-2334-14-124

223 Yusuf I, Arzai AH, Haruna M *et al.* Detection of multi drug resistant bacteria in major hospitals in Kano, North-West, Nigeria. *Braz J Microbiol* 2014;**45**:791–8.

224 Alabi AS, Frielinghaus L, Kaba H *et al.* Retrospective analysis of antimicrobial resistance and bacterial spectrum of infection in Gabon, Central Africa. *BMC Infect Dis* 2013;**13**:455. doi:10.1186/1471-2334-13-455

225 Ibrahim ME, Bilal NE, Magzoub MA *et al.* Prevalence of Extended-spectrum beta-Lactamases-producing *Escherichia coli* from Hospitals in Khartoum State, Sudan. *Oman Med J* 2013;**28**:116–20. doi:10.5001/omj.2013.30

226 Seni J, Peirano G, Okon KO *et al.* The population structure of clinical extra-intestinal *Escherichia coli* in a teaching hospital from Nigeria. *Diagn Microbiol Infect Dis* 2018;**92**:46–9. doi:10.1016/j.diagmicrobio.2018.04.001

227 Obeng-Nkrumah N, Twum-Danso K, Krogfelt KA *et al.* High levels of extended-spectrum beta-lactamases in a major teaching hospital in Ghana: the need for regular monitoring and evaluation of antibiotic resistance. *Am J Trop Med Hyg* 2013;**89**:960–4. doi:10.4269/ajtmh.12-0642

228 Raji MA, Jamal W, Ojemhen O *et al.* Point-surveillance of antibiotic resistance in Enterobacteriaceae isolates from patients in a Lagos Teaching Hospital, Nigeria. *J Infect Public Health* 2013;**6**:431–7. doi:10.1016/j.jiph.2013.05.002

229 Meeren BT van der, Chhaganlal KD, Pfeiffer A *et al.* Extremely high prevalence of

- multi-resistance among uropathogens from hospitalised children in Beira, Mozambique. *S Afr Med J* 2013;**103**:382–6. doi:10.7196/samj.5941
- 230 Idowu OJ, Onipede AO, Orimolade AE *et al.* Extended-spectrum Beta-lactamase Orthopedic Wound Infections in Nigeria. *J Glob Infect Dis* 2011;**3**:211–5. doi:10.4103/0974-777x.83524
- 231 Moyo SJ, Aboud S, Kasubi M *et al.* Antimicrobial resistance among producers and non-producers of extended spectrum beta-lactamases in urinary isolates at a tertiary Hospital in Tanzania. *BMC Res Notes* 2010;**3**:348. doi:10.1186/1756-0500-3-348
- 232 Bercion R, Mossoro-Kpinde D, Manirakiza A *et al.* Increasing prevalence of antimicrobial resistance among Enterobacteriaceae uropathogens in Bangui, Central African Republic. *J Infect Dev Ctries* 2009;**3**:187–90. <https://jidc.org/index.php/journal/article/download/19759473/21>
- 233 Mshana SE, Kamugisha E, Mirambo M *et al.* Prevalence of multiresistant gram-negative organisms in a tertiary hospital in Mwanza, Tanzania. *BMC Res Notes* 2009;**2**:49. doi:10.1186/1756-0500-2-49
- 234 Sire JM, Nabeth P, Perrier-Gros-Claude JD *et al.* Antimicrobial resistance in outpatient Escherichia coli urinary isolates in Dakar, Senegal. *J Infect Dev Ctries* 2007;**1**:263–8.
- 235 Blomberg B, Jureen R, Manji KP *et al.* High rate of fatal cases of pediatric septicemia caused by gram-negative bacteria with extended-spectrum beta-lactamases in Dar es Salaam, Tanzania. *J Clin Microbiol* 2005;**43**:745–9. doi:10.1128/jcm.43.2.745-749.2005
- 236 Gangoue-Pieboji J, Bedenic B, Koulla-Shiro S *et al.* Extended-spectrum-beta-lactamase-producing Enterobacteriaceae in Yaounde, Cameroon. *J Clin Microbiol* 2005;**43**:3273–7. doi:10.1128/jcm.43.7.3273-3277.2005
- 237 Zeynudin A, Pritsch M, Schubert S *et al.* Prevalence and antibiotic susceptibility pattern of CTX-M type extended-spectrum beta-lactamases among clinical isolates of gram-negative bacilli in Jimma, Ethiopia. *BMC Infect Dis* 2018;**18**:524. doi:10.1186/s12879-018-3436-7
- 238 Ndugulile F, Jureen R, Harthug S *et al.* Extended spectrum beta-lactamases among Gram-negative bacteria of nosocomial origin from an intensive care unit of a tertiary health facility in Tanzania. *BMC Infect Dis* 2005;**5**:86. doi:10.1186/1471-2334-5-86
- 239 Dromigny JA, Ndoye B, Macondo EA *et al.* Increasing prevalence of antimicrobial resistance among Enterobacteriaceae uropathogens in Dakar, Senegal: A multicenter study. *Diagn Microbiol Infect Dis* 2003;**47**:595–600. doi:10.1016/S0732-8893(03)00155-X
- 240 Dromigny JA, Nabeth P, Perrier Gros Claude JD. Distribution and susceptibility of bacterial urinary tract infections in Dakar, Senegal. *Int J Antimicrob Agents* 2002;**20**:339–47.

- 241 Ampaire L, Nduhura E, Wewedru I. Phenotypic prevalence of extended spectrum beta-lactamases among enterobacteriaceae isolated at Mulago National Referral Hospital: Uganda. *BMC Res Notes* 2017;**10**:448. doi:10.1186/s13104-017-2786-3
- 242 Andrew B, Kagirita A, Bazira J. Prevalence of Extended-Spectrum Beta-Lactamases-Producing Microorganisms in Patients Admitted at KRRH, Southwestern Uganda. *Int J Microbiol* 2017;**2017**:3183076. doi:10.1155/2017/3183076
- 243 Archary M, Adler H, La Russa P *et al.* Bacterial infections in HIV-infected children admitted with severe acute malnutrition in Durban, South Africa. *Paediatr Int Child Health* 2017;**37**:6–13. doi:10.1080/20469047.2016.1198561
- 244 Henson SP, Boinett CJ, Ellington MJ *et al.* Molecular epidemiology of *Klebsiella pneumoniae* invasive infections over a decade at Kilifi County Hospital in Kenya. *Int J Med Microbiol* 2017;**307**:422–9. doi:10.1016/j.ijmm.2017.07.006
- 245 Chirindze LM, Zimba TF, Sekyere JO *et al.* Faecal colonization of *E. coli* and *Klebsiella* spp. producing extended-spectrum beta-lactamases and plasmid-mediated AmpC in Mozambican university students. *BMC Infect Dis* 2018;**18**:244. doi:10.1186/s12879-018-3154-1
- 246 Founou RC, Founou LL, Essack SY. Extended spectrum beta-lactamase mediated resistance in carriage and clinical gram-negative ESKAPE bacteria: a comparative study between a district and tertiary hospital in South Africa. *Antimicrob Resist Infect Control* 2018;**7**:134. doi:10.1186/s13756-018-0423-0
- 247 Magwenzi MT, Gudza-Mugabe M, Mujuru HA *et al.* Carriage of antibiotic-resistant Enterobacteriaceae in hospitalised children in tertiary hospitals in Harare, Zimbabwe. *Antimicrob Resist Infect Control* 2017;**6**:10. doi:10.1186/s13756-016-0155-y
- 248 Moremi N, Claus H, Vogel U *et al.* Faecal carriage of CTX-M extended-spectrum beta-lactamase-producing Enterobacteriaceae among street children dwelling in Mwanza city, Tanzania. *PLoS One* 2017;**12**:e0184592. doi:10.1371/journal.pone.0184592
- 249 Wilmore SMS, Kranzer K, Williams A *et al.* Carriage of extended-spectrum beta-lactamase-producing Enterobacteriaceae in HIV-infected children in Zimbabwe. *J Med Microbiol* 2017;**66**:609–15. doi:10.1099/jmm.0.000474
- 250 Farra A, Frank T, Tondeur L *et al.* High rate of faecal carriage of extended-spectrum beta-lactamase-producing Enterobacteriaceae in healthy children in Bangui, Central African Republic. *Clin Microbiol Infect* 2016;**22**:891.e1–4. doi:10.1016/j.cmi.2016.07.001
- 251 Desta K, Woldeamanuel Y, Azazh A *et al.* High Gastrointestinal Colonization Rate with Extended-Spectrum beta-Lactamase-Producing Enterobacteriaceae in Hospitalized Patients: Emergence of Carbapenemase-Producing *K. pneumoniae* in Ethiopia. *PLoS One*

2016;**11**:e0161685. doi:10.1371/journal.pone.0161685

252 Djuikoue IC, Woerther PL, Toukam M *et al.* Intestinal carriage of Extended Spectrum Beta-Lactamase producing *E. coli* in women with urinary tract infections, Cameroon. *J Infect Dev Ctries* 2016;**10**:1135–9. doi:10.3855/jidc.7616

253 Mshana SE, Falgenhauer L, Mirambo MM *et al.* Predictors of blaCTX-M-15 in varieties of *Escherichia coli* genotypes from humans in community settings in Mwanza, Tanzania. *BMC Infect Dis* 2016;**16**:187. doi:10.1186/s12879-016-1527-x

254 Ribeiro TG, Novais Â, Peixe L *et al.* Atypical epidemiology of CTX-M-15 among Enterobacteriaceae from a high diversity of non-clinical niches in Angola. *Journal of Antimicrobial Chemotherapy* 2016;**71**:1169–73. doi:10.1093/jac/dkv489

255 Tellevik MG, Blomberg B, Kommedal O *et al.* High Prevalence of Faecal Carriage of ESBL-Producing Enterobacteriaceae among Children in Dar es Salaam, Tanzania. *PLoS One* 2016;**11**:e0168024. doi:10.1371/journal.pone.0168024

256 Chereau F, Herindrainy P, Garin B *et al.* Colonization of extended-spectrum-beta-lactamase- and NDM-1-producing Enterobacteriaceae among pregnant women in the community in a low-income country: a potential reservoir for transmission of multiresistant Enterobacteriaceae to neonates. *Antimicrob Agents Chemother* 2015;**59**:3652–5. doi:10.1128/aac.00029-15

257 Herindrainy P, Rabenandrasana MAN, Andrianirina ZZ *et al.* Acquisition of extended spectrum beta-lactamase-producing enterobacteriaceae in neonates: A community based cohort in Madagascar. *PLoS One* 2018;**13**:e0193325. doi:10.1371/journal.pone.0193325

258 Micheel V, Hogan B, Rakotoarivelo RA *et al.* Identification of nasal colonization with beta-lactamase-producing Enterobacteriaceae in patients, health care workers and students in Madagascar. *Eur J Microbiol Immunol (Bp)* 2015;**5**:116–25. doi:10.1556/eujmi-d-15-00001

259 Nelson E, Kayega J, Seni J *et al.* Evaluation of existence and transmission of extended spectrum beta lactamase producing bacteria from post-delivery women to neonates at Bugando Medical Center, Mwanza-Tanzania. *BMC Res Notes* 2014;**7**:279. doi:10.1186/1756-0500-7-279

260 Lonchel CM, Melin P, Gangoue-Pieboji J *et al.* Extended-spectrum beta-lactamase-producing Enterobacteriaceae in Cameroonian hospitals. *Eur J Clin Microbiol Infect Dis* 2013;**32**:79–87. doi:10.1007/s10096-012-1717-4

261 Magoue CL, Melin P, Gangoue-Pieboji J *et al.* Prevalence and spread of extended-spectrum beta-lactamase-producing Enterobacteriaceae in Ngaoundere, Cameroon. *Clin Microbiol Infect* 2013;**19**:E416–20. doi:10.1111/1469-0691.12239

- 262 Schaumburg F, Alabi A, Kokou C *et al.* High burden of extended-spectrum beta-lactamase-producing Enterobacteriaceae in Gabon. *J Antimicrob Chemother* 2013;**68**:2140–3. doi:10.1093/jac/dkt164
- 263 Albrechtova K, Dolejska M, Cizek A *et al.* Dogs of nomadic pastoralists in northern Kenya are reservoirs of plasmid-mediated cephalosporin- and quinolone-resistant *Escherichia coli*, including pandemic clone B2-O25-ST131. *Antimicrob Agents Chemother* 2012;**56**:4013–7. doi:10.1128/aac.05859-11
- 264 Isendahl J, Turlej-Rogacka A, Manjuba C *et al.* Fecal carriage of ESBL-producing *E. coli* and *K. pneumoniae* in children in Guinea-Bissau: a hospital-based cross-sectional study. *PLoS One* 2012;**7**:e51981. doi:10.1371/journal.pone.0051981
- 265 Lonchel CM, Meex C, Gangoué-Piéboji J *et al.* Proportion of extended-spectrum  $\beta$ -lactamase-producing Enterobacteriaceae in community setting in Ngaoundere, Cameroon. *BMC Infect Dis* 2012;**12**. doi:10.1186/1471-2334-12-53
- 266 Herindrainy P, Randrianirina F, Ratovoson R *et al.* Rectal carriage of extended-spectrum beta-lactamase-producing gram-negative bacilli in community settings in Madagascar. *PLoS One* 2011;**6**:e22738. doi:10.1371/journal.pone.0022738
- 267 Woerther PL, Angebault C, Jacquier H *et al.* Massive increase, spread, and exchange of extended spectrum beta-lactamase-encoding genes among intestinal Enterobacteriaceae in hospitalized children with severe acute malnutrition in Niger. *Clin Infect Dis* 2011;**53**:677–85. doi:10.1093/cid/cir522
- 268 Katakweba AAS, Muhairwa AP, Lupindu AM *et al.* First Report on a Randomized Investigation of Antimicrobial Resistance in Fecal Indicator Bacteria from Livestock, Poultry, and Humans in Tanzania. *Microbial Drug Resistance* 2018;**24**:260–8. doi:10.1089/mdr.2016.0297
- 269 Andriatahina T, Randrianirina F, Hariniana ER *et al.* High prevalence of fecal carriage of extended-spectrum beta-lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae* in a pediatric unit in Madagascar. *BMC Infect Dis* 2010;**10**:204. doi:10.1186/1471-2334-10-204
- 270 Ruppe E, Woerther PL, Diop A *et al.* Carriage of CTX-M-15-producing *Escherichia coli* isolates among children living in a remote village in Senegal. *Antimicrob Agents Chemother* 2009;**53**:3135–7. doi:10.1128/aac.00139-09
- 271 Tande D, Jallot N, Bougoudogo F *et al.* Extended-spectrum beta-lactamase-producing Enterobacteriaceae in a Malian orphanage. *Emerg Infect Dis* 2009;**15**:472–4. doi:10.3201/eid1503.071637
- 272 Marando R, Seni J, Mirambo MM *et al.* Predictors of the extended-spectrum-beta lactamases producing Enterobacteriaceae neonatal sepsis at a tertiary hospital, Tanzania.



*Int J Med Microbiol* 2018;**308**:803–11. doi:10.1016/j.ijmm.2018.06.012

273 Moremi N, Claus H, Rutta L *et al.* High carriage rate of extended-spectrum beta-lactamase-producing Enterobacteriaceae among patients admitted for surgery in Tanzanian hospitals with a low rate of endogenous surgical site infections. *J Hosp Infect* 2018;**100**:47–53. doi:10.1016/j.jhin.2018.05.017

274 Nikiema Pessinaba C, Landoh DE, Dossim S *et al.* Screening for extended-spectrum beta-lactamase-producing Enterobacteriaceae intestinal carriage among children aged under five in Lome, Togo. *Med Mal Infect* 2018;**48**:551–4. doi:10.1016/j.medmal.2018.07.004

275 Sanneh B, Kebbeh A, Jallow HS *et al.* Prevalence and risk factors for faecal carriage of Extended Spectrum beta-lactamase producing Enterobacteriaceae among food handlers in lower basic schools in West Coast Region of The Gambia. *PLoS One* 2018;**13**:e0200894. doi:10.1371/journal.pone.0200894

276 Stanley IJ, Kajumbula H, Bazira J *et al.* Multidrug resistance among Escherichia coli and Klebsiella pneumoniae carried in the gut of out-patients from pastoralist communities of Kasese district, Uganda. *PLoS One* 2018;**13**. doi:10.1371/journal.pone.0200093

277 Kurz MS, Bayingana C, Ndoli JM *et al.* Intense pre-admission carriage and further acquisition of ESBL-producing Enterobacteriaceae among patients and their caregivers in a tertiary hospital in Rwanda. *Trop Med Int Health* 2017;**22**:210–20. doi:10.1111/tmi.12824

278 Gray KJ, Wilson LK, Phiri A *et al.* Identification and characterization of ceftriaxone resistance and extended-spectrum  $\beta$ -lactamases in Malawian bacteraemic Enterobacteriaceae. *Journal of Antimicrobial Chemotherapy* 2006;**57**:661–5. doi:10.1093/jac/dkl037

279 Musicha P, Feasey NA, Cain AK *et al.* Genomic landscape of extended-spectrum  $\beta$ -lactamase resistance in Escherichia coli from an urban African setting. *Journal of Antimicrobial Chemotherapy* 2017;**72**:1602–9. doi:10.1093/jac/dkx058

280 Musicha P, Msefula CL, Mather AE *et al.* Genomic analysis of Klebsiella pneumoniae isolates from Malawi reveals acquisition of multiple ESBL determinants across diverse lineages. *Journal of Antimicrobial Chemotherapy* 2019;**74**:1223–32. doi:10.1093/jac/dkz032

281 National Statistics Office of Malawi. Malawi population projections. [www.nsomalawi.mw](http://www.nsomalawi.mw) (accessed 10 Feb 2019).

282 World Bank. World Bank Country and Lending Groups. <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups> (accessed 28 Feb 2019).

283 United Nations Development Programme. Human Development Indices and Indicators: 2018 Statistical Update. 2018.

- 284 UNAIDS. Malawi Country Profile. <http://aidsinfo.unaids.org/> (accessed 28 Feb 2019).
- 285 World Health Organisation. Global tuberculosis report. Geneva: 2018.
- 286 British Society for Antimicrobial Chemotherapy. BSAC Methods for Antimicrobial Susceptibility Testing. 2013.
- 287 Ministry of Health. Government of Malawi. Malawi HIV Testing Service Guidelines. 2015.
- 288 Barrow GI, Cowan ST, Steel KJ *et al.* *Cowan and Steel's manual for the identification of medical bacteria*. Cambridge University Press 1993.
- 289 World Health Organisation. The use of lateral flow urine lipoarabinomannan assay (LF-LAM) for the diagnosis and screening of active tuberculosis in people living with HIV. Geneva: 2015.
- 290 Sutherland A, Rattray J. SCIEH Weekly Report HEALTH PROTECTION SCOTLAND Supplement to SCIEH Weekly Report Summary. 2004. <http://www.show.scot.nhs.uk/scieh/>
- 291 Hartung C, Lerer A, Anokwa Y *et al.* Open data kit: tools to build information services for developing regions. In: *Proceedings of the 4th acm/ieee international conference on information and communication technologies and development - ictd '10*. New York, New York, USA: 2010. doi:10.1145/2369220.2369236
- 292 Larsson J. eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses. 2019.<https://cran.r-project.org/package=eulerr>
- 293 Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 2017;**33**:2938–40. doi:10.1093/bioinformatics/btx364
- 294 Therneau TM. Survival: A Package for Survival Analysis in S. 2015.<https://cran.r-project.org/package=survival>
- 295 Jelsma J, Hansen K, De Weerd W *et al.* How do Zimbabweans value health states? *Population health metrics* 2003;**1**:11. doi:10.1186/1478-7954-1-11
- 296 Huson MAM, Kalkman R, Stolp SM *et al.* The impact of HIV on presentation and outcome of bacterial sepsis and other causes of acute febrile illness in Gabon. *Infection* 2015;**43**:443–51. doi:10.1007/s15010-015-0753-2
- 297 Moore CC, Hazard R, Saulters KJ *et al.* Derivation and validation of a universal vital assessment (UVA) score: a tool for predicting mortality in adult hospitalised patients in sub-Saharan Africa. *BMJ Global Health* 2017;**2**:e000344. doi:10.1136/bmjgh-2017-000344
- 298 Feasey NA, Masesa C, Jassi C *et al.* Three Epidemics of Invasive Multidrug-Resistant Salmonella Bloodstream Infection in Blantyre, Malawi, 1998-2014. *Clinical infectious*

*diseases : an official publication of the Infectious Diseases Society of America* 2015;**61** Suppl 4:S363–71. doi:10.1093/cid/civ691

299 Reddy EA, Shaw AV, Crump JA. Community-acquired bloodstream infections in Africa: a systematic review and meta-analysis. *Lancet Infect Dis* 2010;**10**:417–32. doi:10.1016/s1473-3099(10)70072-4

300 Doolan DL, Dobaño C, Baird JK. Acquired immunity to malaria. *Clinical microbiology reviews* 2009;**22**:13–36. doi:10.1128/CMR.00025-08

301 Gaskell KM, Rothe C, Gnanadurai R *et al.* A Prospective Study of Mortality from Cryptococcal Meningitis following Treatment Induction with 1200mg Oral Fluconazole in Blantyre, Malawi. *PLoS ONE* 2014;**9**:e110285. doi:10.1371/journal.pone.0110285

302 Levy MM, Evans LE, Rhodes A. The Surviving Sepsis Campaign Bundle: 2018 update. *Intensive Care Medicine* 2018;**44**:925–8. doi:10.1007/s00134-018-5085-0

303 Hazard RH, Kagina P, Kitayimbwa R *et al.* Effect of Empiric Anti-Mycobacterium tuberculosis Therapy on Survival Among Human Immunodeficiency Virus-Infected Adults Admitted With Sepsis to a Regional Referral Hospital in Uganda. *Open Forum Infectious Diseases* 2019;**6**. doi:10.1093/ofid/ofz140

304 Hosseinipour MC, Bisson GP, Miyahara S *et al.* Empirical tuberculosis therapy versus isoniazid in adult outpatients with advanced HIV initiating antiretroviral therapy (REMEMBER): a multicountry open-label randomised controlled trial. *The Lancet* 2016;**387**:1198–209. doi:10.1016/S0140-6736(16)00546-8

305 Blanc F-X, Badje2 AD, Bonnet M *et al.* Systematic vs test-guided tuberculosis treatment: data of the STASIS randomized trial. In: *Conference on retroviruses and opportunistic infections*. 2018.

306 Grant AD, Charalambous S, Tlali M *et al.* Algorithm-guided empirical tuberculosis treatment for people with advanced HIV (TB Fast Track): an open-label, cluster-randomised trial. *The Lancet HIV* 2019;**Online Fir**. doi:10.1016/S2352-3018(19)30266-8

307 Cockerill FR, Wilson JW, Vetter EA *et al.* Optimal Testing Parameters for Blood Cultures. *Clinical Infectious Diseases* 2004;**38**:1724–30. doi:10.1086/421087

308 Barr DA, Kerkhoff AD, Schutz C *et al.* HIV-Associated Mycobacterium tuberculosis Bloodstream Infection Is Underdiagnosed by Single Blood Culture. *Journal of clinical microbiology* 2018;**56**:e01914–17. doi:10.1128/JCM.01914-17

309 Harrell FE. *Regression modeling strategies : with applications to linear models, logistic regression, and survival analysis*. Springer 2001.

310 Westreich D, Greenland S. The Table 2 Fallacy: Presenting and Interpreting Con-

founder and Modifier Coefficients. *American Journal of Epidemiology* 2013;**177**:292–8. doi:10.1093/aje/kws412

311 Hastie T, Tibshirani R, Friedman JH(H. *The elements of statistical learning : data mining, inference, and prediction*. Springer 2009.

312 Tibshirani R. Regression Shrinkage and Selection via the Lasso. 1996;**58**:267–88. doi:10.2307/2346178

313 Lê S, Josse J, Husson F. FactoMineR : An R Package for Multivariate Analysis. *Journal of Statistical Software* 2008;**25**:1–18. doi:10.18637/jss.v025.i01

314 Gelman A, Hill J. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press 2007.

315 Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ, USA:: John Wiley & Sons, Inc. 1987. doi:10.1002/9780470316696

316 Gelman A, Carlin JB, Stern HS *et al*. *Bayesian data analysis*. 3rd ed. Chapman; Hall/CRC 2004.

317 Pearl J, Glymour M, Jewell NP. *Causal inference in statistics : a primer*. John Wiley & Sons Ltd 2016.

318 Textor J, Zander B van der, Gilthorpe MS *et al*. Robust causal inference using directed acyclic graphs: the R package ‘dagitty’. *International Journal of Epidemiology* 2017;**45**:dyw341. doi:10.1093/ije/dyw341

319 Bürkner P-C. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* 2017;**80**:1–28. doi:10.18637/jss.v080.i01

320 Carpenter B, Gelman A, Hoffman MD *et al*. Stan: A Probabilistic Programming Language. *Journal of Statistical Software* 2017;**76**:1–32. doi:10.18637/jss.v076.i01

321 Buuren S van, Groothuis-Oudshoorn K. mice : Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* 2011;**45**:1–67. doi:10.18637/jss.v045.i03

322 Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing* 2017;**27**:1413–32. doi:10.1007/s11222-016-9696-4

323 Cleveland WS, Devlin SJ. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association* 1988;**83**:596–610. doi:10.1080/01621459.1988.10478639

324 Ho DE, Imai K, King G *et al*. MatchIt : Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software* 2011;**42**:1–28. doi:10.18637/jss.v042.i08

- 325 Feasey NA, Gaskell K, Wong V *et al.* Rapid Emergence of Multidrug Resistant, H58-Lineage Salmonella Typhi in Blantyre, Malawi. *PLOS Neglected Tropical Diseases* 2015;**9**:e0003748. doi:10.1371/journal.pntd.0003748
- 326 Seymour CW, Kennedy JN, Wang S *et al.* Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. *JAMA* 2019;**321**:2003. doi:10.1001/jama.2019.5791
- 327 Broger T, Sossen B, Toit E du *et al.* Novel lipoarabinomannan point-of-care tuberculosis test for people with HIV: a diagnostic accuracy study. *The Lancet Infectious Diseases* 2019;**19**:852–61. doi:10.1016/S1473-3099(19)30001-5
- 328 Dillon SM, Frank DN, Wilson CC. The gut microbiome and HIV-1 pathogenesis: a two-way street. *AIDS* 2016;**30**:2737–51. doi:10.1097/QAD.0000000000001289
- 329 UNIADS. UNIADS country profile: Malawi. 2018.
- 330 Malawi Ministry of Health. Malawi Guidelines for Clinical Management of HIV in Children and Adults (Third Edition). 2016.
- 331 Ministry of Health. Government of Malawi. Malawi standard treatment guidelines (MSTG) 5th edition 2015. 2015.
- 332 Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* 2014;**15**:R46. doi:10.1186/gb-2014-15-3-r46
- 333 Bankevich A, Nurk S, Antipov D *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* 2012;**19**:455–77. doi:10.1089/cmb.2012.0021
- 334 Gurevich A, Saveliev V, Vyahhi N *et al.* QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;**29**:1072–5. doi:10.1093/bioinformatics/btt086
- 335 Parks DH, Imelfort M, Skennerton CT *et al.* CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research* 2015;**25**:1043–55. doi:10.1101/gr.186072.114
- 336 Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;**30**:2068–9. doi:10.1093/bioinformatics/btu153
- 337 Page AJ, Cummins CA, Hunt M *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;**31**:3691–3. doi:10.1093/bioinformatics/btv421
- 338 Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* 2013;**30**:772–80. doi:10.1093/molbev/mst010

- 339 Page AJ, Taylor B, Delaney AJ *et al.* SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial Genomics* 2016;**2**. doi:10.1099/mgen.0.000056
- 340 Nguyen L-T, Schmidt HA, Haeseler A von *et al.* IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* 2015;**32**:268–74. doi:10.1093/molbev/msu300
- 341 Yu G, Smith DK, Zhu H *et al.* Ggtree : An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* 2017;**8**:28–36. doi:10.1111/2041-210X.12628
- 342 Hunt M, Mather AE, Sánchez-Busó L *et al.* ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microbial genomics* 2017;**3**:e000131. doi:10.1099/mgen.0.000131
- 343 Inouye M, Dashnow H, Raven L-A *et al.* SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Medicine* 2014;**6**:90. doi:10.1186/s13073-014-0090-6
- 344 Carattoli A, Zankari E, García-Fernández A *et al.* In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrobial agents and chemotherapy* 2014;**58**:3895–903. doi:10.1128/AAC.02412-14
- 345 Clermont O, Christenson JK, Denamur E *et al.* The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environmental Microbiology Reports* 2013;**5**:58–65. doi:10.1111/1758-2229.12019
- 346 Runcharoen C, Raven KE, Reuter S *et al.* Whole genome sequencing of ESBL-producing *Escherichia coli* isolated from patients, farm waste and canals in Thailand. *Genome Medicine* 2017;**9**:81. doi:10.1186/s13073-017-0471-8
- 347 Mentzer A von, Connor TR, Wieler LH *et al.* Identification of enterotoxigenic *Escherichia coli* (ETEC) clades with long-term global distribution. *Nature Genetics* 2014;**46**:1321–6. doi:10.1038/ng.3145
- 348 Ingle DJ, Tauschek M, Edwards DJ *et al.* Evolution of atypical enteropathogenic *E. coli* by repeated acquisition of LEE pathogenicity island variants. *Nature Microbiology* 2016;**1**:15010. doi:10.1038/nmicrobiol.2015.10
- 349 Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal* 2006;1695.<http://igraph.org>
- 350 Dale AP, Woodford N. Extra-intestinal pathogenic *Escherichia coli* (ExPEC): Disease, carriage and clones. *Journal of Infection* 2015;**71**:615–26. doi:10.1016/J.JINF.2015.09.009
- 351 Krishnaraju M, Kamatchi C, Jha AK *et al.* Complete sequencing of an IncX3 plasmid

carrying blaNDM-5 allele reveals an early stage in the dissemination of the blaNDM gene. *Indian journal of medical microbiology* 2015;**33**:30–8. doi:10.4103/0255-0857.148373

352 Robicsek A, Strahilevitz J, Jacoby GA *et al.* Fluoroquinolone-modifying enzyme: a new adaptation of a common aminoglycoside acetyltransferase. *Nature Medicine* 2006;**12**:83–8. doi:10.1038/nm1347

353 Yan L, Liu D, Wang X-H *et al.* Bacterial plasmid-mediated quinolone resistance genes in aquatic environments in China. *Scientific reports* 2017;**7**:40610. doi:10.1038/srep40610

354 Ramirez MS, Tolmasky ME. Aminoglycoside modifying enzymes. *Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy* 2010;**13**:151–71. doi:10.1016/j.drug.2010.08.003

355 Galimand M, Courvalin P, Lambert T. Plasmid-Mediated High-Level Resistance to Aminoglycosides in Enterobacteriaceae Due to 16S rRNA Methylation. *Antimicrobial Agents and Chemotherapy* 2003;**47**:2565. doi:10.1128/AAC.47.8.2565-2571.2003

356 Anantham S, Hall RM. pCERC1, a Small, Globally Disseminated Plasmid Carrying the dfrA14 Cassette in the strA Gene of the sul2-strA-strB Gene Cluster. *Microbial Drug Resistance* 2012;**18**:364–71. doi:10.1089/mdr.2012.0008

357 Roer L, Overballe-Petersen S, Hansen F *et al.* Escherichia coli Sequence Type 410 Is Causing New International High-Risk Clones. *mSphere* 2018;**3**:e00337–18. doi:10.1128/mSphere.00337-18

358 Zong Z, Fenn S, Connor C *et al.* Complete genomic characterization of two Escherichia coli lineages responsible for a cluster of carbapenem-resistant infections in a Chinese hospital. *Journal of Antimicrobial Chemotherapy* 2018;**73**:2340–6. doi:10.1093/jac/dky210

359 Wang Q, Wang X, Wang J *et al.* Phenotypic and Genotypic Characterization of Carbapenem-resistant Enterobacteriaceae : Data From a Longitudinal Large-scale CRE Study in China (2012–2016). *Clinical Infectious Diseases* 2018;**67**:S196–205. doi:10.1093/cid/ciy660

360 Irrgang A, Falgenhauer L, Fischer J *et al.* CTX-M-15-Producing E. coli Isolates from Food Products in Germany Are Mainly Associated with an IncF-Type Plasmid and Belong to Two Predominant Clonal E. coli Lineages. *Frontiers in Microbiology* 2017;**8**:2318. doi:10.3389/fmicb.2017.02318

361 Kallonen T, Brodrick HJ, Harris SR *et al.* Systematic longitudinal survey of invasive Escherichia coli in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. *Genome Research* 2017;**27**:1437–49. doi:10.1101/gr.216606.116

362 Pitout JDD, DeVinney R. Escherichia coli ST131: a multidrug-resistant clone primed

- for global domination. *F1000Research* 2017;**6**:195. doi:10.12688/f1000research.10609.1
- 363 Redgrave LS, Sutton SB, Webber MA *et al.* Fluoroquinolone resistance: mechanisms, impact on bacteria, and role in evolutionary success. *Trends in Microbiology* 2014;**22**:438–45. doi:10.1016/j.tim.2014.04.007
- 364 Ostrer L, Khodursky RF, Johnson JR *et al.* Analysis of mutational patterns in quinolone resistance-determining regions of GyrA and ParC of clinical isolates. *International Journal of Antimicrobial Agents* 2019;**53**:318–24. doi:10.1016/j.ijantimicag.2018.12.004
- 365 Garoff L, Yadav K, Hughes D. Increased expression of Qnr is sufficient to confer clinical resistance to ciprofloxacin in *Escherichia coli*. *Journal of Antimicrobial Chemotherapy* 2018;**73**:348–52. doi:10.1093/jac/dkx375
- 366 Neuert S, Nair S, Day MR *et al.* Prediction of Phenotypic Antimicrobial Resistance Profiles From Whole Genome Sequences of Non-typhoidal *Salmonella enterica*. *Frontiers in microbiology* 2018;**9**:592. doi:10.3389/fmicb.2018.00592
- 367 Williams CT, Musicha P, Feasey NA *et al.* ChloS-HRM, a novel assay to identify chloramphenicol-susceptible *Escherichia coli* and *Klebsiella pneumoniae* in Malawi. *The Journal of antimicrobial chemotherapy* 2019;**74**:1212. doi:10.1093/jac/dky563
- 368 Totsika M, Beatson SA, Sarkar S *et al.* Insights into a Multidrug Resistant *Escherichia coli* Pathogen of the Globally Disseminated ST131 Lineage: Genome Analysis and Virulence Mechanisms. *PLoS ONE* 2011;**6**:e26578. doi:10.1371/JOURNAL.PONE.0026578
- 369 Sekizuka T, Yatsu K, Inamine Y *et al.* Complete Genome Sequence of a blaKPC-2-Positive *Klebsiella pneumoniae* Strain Isolated from the Effluent of an Urban Sewage Treatment Plant in Japan. *mSphere* 2018;**3**. doi:10.1128/mSphere.00314-18
- 370 Cheng L, Connor TR, Siren J *et al.* Hierarchical and Spatially Explicit Clustering of DNA Sequences with BAPS Software. *Molecular Biology and Evolution* 2013;**30**:1224–8. doi:10.1093/molbev/mst028
- 371 Altschul SF, Gish W, Miller W *et al.* Basic local alignment search tool. *Journal of Molecular Biology* 1990;**215**:403–10. doi:10.1016/S0022-2836(05)80360-2
- 372 Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 2006;**22**:1658–9. doi:10.1093/bioinformatics/btl158
- 373 Siguier P, Perochon J, Lestrade L *et al.* ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Research* 2006;**34**:D32–6. doi:10.1093/nar/gkj014
- 374 Stoesser N, Sheppard AE, Moore CE *et al.* Extensive Within-Host Diversity in Fecally Carried Extended-Spectrum-Beta-Lactamase-Producing *Escherichia coli* Isolates: Implications for Transmission Analyses. *Journal of clinical microbiology* 2015;**53**:2122–31.



doi:10.1128/JCM.00378-15

375 Kluytmans-van den Bergh MFQ, Mens SP van, Haverkate MR *et al.* Quantifying Hospital-Acquired Carriage of Extended-Spectrum Beta-Lactamase-Producing Enterobacteriaceae Among Patients in Dutch Hospitals. *Infection Control & Hospital Epidemiology* 2018;**39**:32–9. doi:10.1017/ice.2017.241

376 Haverkate M, Platteel T, Fluit A *et al.* Quantifying within-household transmission of extended-spectrum  $\beta$ -lactamase-producing bacteria. *Clinical Microbiology and Infection* 2017;**23**:46.e1–7. doi:10.1016/j.cmi.2016.08.021

377 Hout A van den. *Multi-state survival models for interval-censored data*. Chapman; Hall/CRC 2016.

378 Longini IM, Clark WS, Byers RH *et al.* Statistical analysis of the stages of HIV infection using a Markov model. *Statistics in medicine* 1989;**8**:831–43.<http://www.ncbi.nlm.nih.gov/pubmed/2772443>

379 Andersen PK. Multistate models in survival analysis: A study of nephropathy and mortality in diabetes. *Statistics in Medicine* 1988;**7**:661–70. doi:10.1002/sim.4780070605

380 Jackson CH, Sharples LD. Hidden Markov models for the onset and progression of bronchiolitis obliterans syndrome in lung transplant recipients. *Statistics in Medicine* 2002;**21**:113–28. doi:10.1002/sim.886

381 Jackson CH. Multi-State Models for Panel Data: The msm package for R. *Journal of Statistical Software* 2011;**38**:1–28. doi:10.18637/jss.v038.i08

382 Marshall G, Jones RH. Multi-state models and diabetic retinopathy. *Statistics in medicine* 1995;**14**:1975–83.<http://www.ncbi.nlm.nih.gov/pubmed/8677398>

383 Hoffman MD, Gelman A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. 2014. <http://mcmc-jags.sourceforge.net>

384 Betancourt M. Diagnosing Biased Inference with Divergences. 2017.[https://mc-stan.org/users/documentation/case-studies/divergences{\\\_}and{\\\_}bias.html](https://mc-stan.org/users/documentation/case-studies/divergences{\_}and{\_}bias.html) (accessed 1 Jun 2019).

385 Soetaert K, Petzoldt T, Setzer RW. Solving Differential Equations in R : Package deSolve. *Journal of Statistical Software* 2010;**33**:1–25. doi:10.18637/jss.v033.i09

386 Lewis DK, Callaghan M, Phiri K *et al.* Prevalence and indicators of HIV and AIDS among adults admitted to medical and surgical wards in Blantyre, Malawi. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 2003;**97**:91–6. doi:10.1016/S0035-9203(03)90035-6

387 Pallega A, Mikkelsen KH, Forslund SK *et al.* Recovery of gut microbiota of healthy adults

following antibiotic exposure. *Nature Microbiology* 2018;**3**:1255–65. doi:10.1038/s41564-018-0257-9

388 Francino MP. Antibiotics and the Human Gut Microbiome: Dysbioses and Accumulation of Resistances. *Frontiers in Microbiology* 2016;**6**:1543. doi:10.3389/fmicb.2015.01543

389 Buffie CG, Pamer EG. Microbiota-mediated colonization resistance against intestinal pathogens. *Nature reviews Immunology* 2013;**13**:790–801. doi:10.1038/nri3535

390 World Health Organisation. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach. Second edition. Geneva: 2016.

391 Anywaine Z, Levin J, Kasirye R *et al.* Discontinuing cotrimoxazole preventive therapy in HIV-infected adults who are stable on antiretroviral treatment in Uganda (COSTOP): A randomised placebo controlled trial. *PLOS ONE* 2018;**13**:e0206907. doi:10.1371/journal.pone.0206907